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 [Draft]

Committee:

Jacek Gwizdka, Supervisor

Soo-Young Rieh

Matthew Lease

Robert Capra

LongSAL: A Longitudinal Search as Learning Study With University Students [Draft]

by

Nilavra Bhattacharya

নীলাভ্র ভট্টাচার্য্য

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 [Draft]

Nilavra Bhattacharya নীলাভ্র ভট্টাচার্য্য, 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. Researchers have expressed concerns over this ephemeral approach, as learning takes place over time, and is not fleeting. In this dissertation, an exploratory longitudinal study was conducted to analyse the long-term searching behaviour of students enrolled in a university course, over the span of a 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.

সারসংক্ষেপ (Abstract in Bengali)

লংস্যাল: ইন্টারনেটে তথ্য অনুসন্ধানের মাধ্যমে শিক্ষার্থীদের জ্ঞানলাভ বোঝার জন্য একটি অনুদৈর্ঘ্য গবেষণা [খসড়া]

নীলাভ্র ভট্টাচার্য্য, PhD TBD
টেক্সাস বিশ্ববিদ্যালয়, অস্টিন, মার্কিন যুক্তরাষ্ট্র, বৈশাখ ১৪৩০ (মে ২০২৩)

সুপারভাইজার: ইয়াৎসেক গুইজদকা

আমার পিএইচডি এডভাইসর পরামর্শ দিয়েছেন যে এই থিসিস এর একটি বাংলা ভাসায় সারসংক্ষেপ ও লিখতে। উনি ওনার থিসিস এর সারসংক্ষেপ ইংরেজি এবং পোলিশ দুই ভাষায় লিখেছিলেন। থিসিসের প্রথম খসড়া জমা দেওয়ার পরে বাংলা সারসংক্ষেপ টি আরও বিস্তারিতভাবে লেখা হবে। বাংলা সংস্করণ এর উৎসর্গ: দিদার জন্য (tie back to PhD Statement of Purpose)

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., 2018, 2019; 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, 2001a; 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

To address the gaps in our knowledge of how information searching influences students’ learning process over time, this dissertation conducted 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 aimed at:

- situating students as learners in their naturalistic contexts, and characterized by their individual differences
- measuring students’ information search and information use behaviour over time

- correlating 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 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" (anonymous reviewer for Bhattacharya, 2021).

1.4 Outline

This dissertation 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

This first chapter on background literature discusses relevant concepts from the disciplines of Education and Learning Sciences. First, we introduce some relevant terminology, and the concepts of deep or meaningful learning. Then we discuss several research backed principles that have been shown to lead to meaningful learning. Next, we discuss how learning, sensemaking, and searching for information are related, and how modern technologies provide affordances for new forms of learning and knowledge work in the 21st century. We also discuss some concepts about individual differences of learners as well as techniques that can promote better learning. In the last section, we state what implications these findings have for shaping the longitudinal study in this dissertation.

2.1 Terminology

The Webster dictionary[1] defines **knowledge** in two ways. The first definition is "the range of one's information or understanding". Vakkari (2016) says it is "the totality what a person knows, that is, a **personal knowledge** or **belief system**. It may include both justified, true beliefs and less justified, not so true beliefs, which the person more or less thinks hold true." Webster's

[1] https://www.merriam-webster.com/dictionary/knowledge

second definition of knowledge is "the sum of what is known: the body of truth, information, and principles acquired by humankind". We can regard this as **universal knowledge**.

Learning is a *process*, that leads to a *change* in (personal) knowledge, beliefs, behaviours, and attitudes (Ambrose et al., 2010). Thus, learning always aims to increase one's personal knowledge, and can often draw from the body of universal knowledge. In some cases, the change in personal knowledge can also lead to change in universal knowledge, such as when new discoveries are made, or new philosophies are proposed. Human learning is an innate capacity. It is longitudinal and unfolds over time. Learning is lifelong and life-wide, and has a lasting impact on how humans think and act (Ambrose et al., 2010; Kalantzis & Cope, 2012). Learning can be informal or formal. **Informal learning** is the casual learning taking place in everyday life, and is incidental to the everyday life experience. **Formal learning** is the deliberate, conscious, systematic, and explicit acquiring of knowledge (Kalantzis & Cope, 2012).

Education is a form of formal learning. It is the systematic acquiring of knowledge. In today's world, the institutions of education are formally constructed places (classrooms), times (of the day and of life) and social relations (teachers and students); for instance, schools, colleges, and universities. The scientific discipline of Education concerns itself with the systematic investigation of the ways in which humans know and learn. It is the science of "coming to know" (Kalantzis & Cope, 2012).

Pedagogy describes small sequences of learner activities that promote learning in educational settings (Kalantzis & Cope, 2012). Traditional approaches to (classroom) pedagogy, especially the *didactic pedagogy*, primarily involves a teacher telling, and a learner listening. The teacher is in command of the knowledge, and their mission is to transmit this knowledge to the learners, in a one-way flow. It is hoped that the learners will dutifully absorb the knowledge laid before them by the teacher. The balance of agency weighs heavily towards the teacher. "There is a special focus on long-term memory, or retention, measurable by the ritual of closed-book, summative examination" (Cope & Kalantzis, 2017).

Cognitive scientists had discovered that learners retain material better, and are able to generalize and apply it to a broader range of contexts, when they learn **deep knowledge** rather

Learning Knowledge Deeply *Findings from Cognitive Science => Reflexive Pedagogy*	**Traditional Classroom Practices** *Didactic Pedagogy => Instructionism / Surface Learning*
Knowledge Integration and Sensemaking:	
• Learners relate new ideas and concepts to previous knowledge and experience • Learners integrate their knowledge into interrelated conceptual systems • Learners look for patterns and underlying principles	• Learners treat course material as unrelated to what they already know • Learners treat course material as disconnected bits of knowledge • Learners memorize facts and carry out procedures without understanding how or why
Active Knowledge Making and Multiliteracy	
• Learners understand the process of dialogue through which knowledge is created, and they examine the logic of an argument critically	• Learners treat facts and procedures as static knowledge, handed down from an all-knowing authority
• Learners are the also knowledge producers, and discerning knowledge discoverers / navigators	• Learner is the knowledge consumer, with passive knowledge acquisition and memorization
• <u>Multiliteracies</u>: learners interact with new forms of media; they consume and produce multimodal knowledge artefacts (images, videos, presentations, software, etc.)	• Academic literacies: learners interact with traditional textbooks, assignments, and tests
Metacognition and Self-regulation:	
• Learners reflect on their own understanding, and their own process of learning • Thinking about thinking, critical self-reflection on knowledge processes and disciplinary practices	• Learners memorize without reflecting on the purpose, or on their own learning strategies. • Focus on facts to be remembered, theories to be correctly applied.

Figure 2.1: Deep learning (of the human kind) versus traditional (also often online) classroom practices. Compiled from Cope & Kalantzis (2017) and Sawyer (2005).

than **surface knowledge**, and when they learn how to use that knowledge in real-world social and practical settings (Sawyer, 2005). Deep learning [2] takes place when "the learner chooses conscientiously to integrate new knowledge to knowledge that the learner already possesses" and involves "substantive, non-arbitrary incorporations of concepts into cognitive structure" (Novak, 2002, p. 549) and may eventually lead to the development of transferable knowledge and skills. A parallel terminology for deep learning (Marton & Säaljö, 1976; Marton & Säljö, 1976) is **meaningful learning** (Ausubel et al., 1968; Novak, 2002), and they are often contrasted with *surface learning* or *rote learning*. Figure 2.1 discusses some more details on deep or meaningful learning, and the limitations of traditional classroom practices to promote deep learning. Figure 2.2 describes (using a concept map) how meaningful learning can be achieved and sustained,

[2] of the human kind

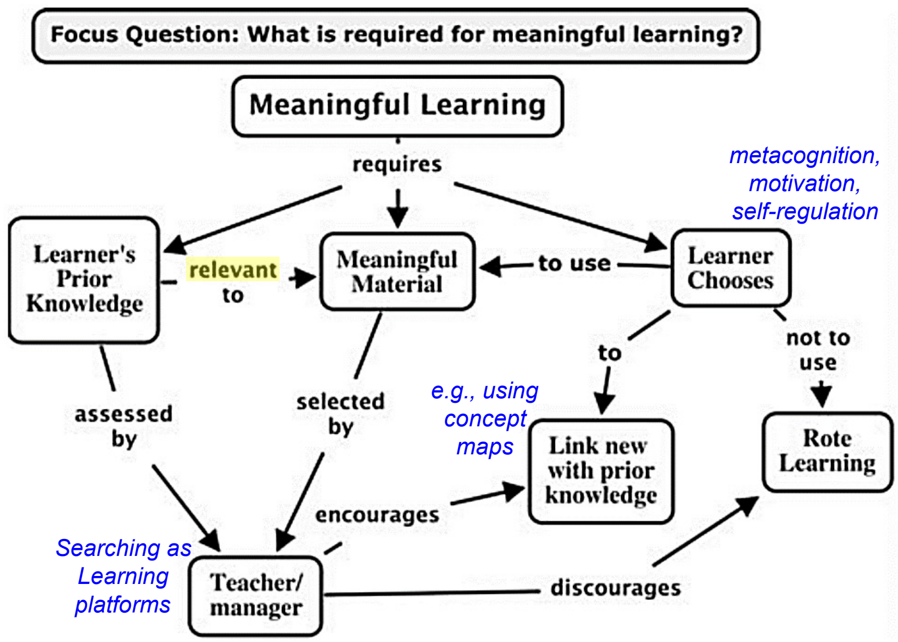


Figure 2.2: Meaningful learning (aka deep learning) as explained by Novak (2010, fig. 5.3) (annotations our own).

and our annotations highlight how Search-as-learning systems can foster the same.

2.2 Principles of Meaningful Learning

Ambrose et al. (2010) have proposed several principles of (student) learning that lead to creation of deeper knowledge in learners, and help educators understand why certain teaching approaches may help or hinder learning. These principles are based on research and literature from a range of disciplines in psychology, education, and anthropology, and the authors claim they are domain independent, experience independent, and cross-culturally relevant.

1. Students' **prior knowledge** can help or hinder learning.
2. How students **organize knowledge** influences how they learn and apply what they know.
3. Students' **motivation** determines, directs, and sustains what they do to learn.
4. Goal-directed practice coupled with **targeted feedback** enhances the quality of students' learning.

5. Students' current level of development interacts with the social, emotional, and intellectual **context** around the student to impact learning.
6. To become **self-directed** learners, students must learn to **monitor and adjust** their approaches to learning.

In line with the above, the US National Research Council identified several key principles about **experts' knowledge** (National Research Council, 2000), that illustrate the outcome of successful learning:

1. Experts notice features and **meaningful patterns** of information that are not noticed by novices.

2. Experts have acquired a great deal of content knowledge that is **organized** in ways that reflect a deep understanding of their subject matter.

3. Experts' knowledge cannot be reduced to sets of isolated facts or propositions but, instead, reflects contexts of **applicability**: that is, the knowledge is 'conditionalized' on a set of circumstances.

4. Experts are able to **flexibly retrieve** important aspects of their knowledge with little attentional effort.

5. Though experts know their disciplines thoroughly, this does not guarantee that they are able to teach others.

6. Experts have varying levels of flexibility in their approach to new situations.

The principles of learning illustrate that both the *context* of learning, and the *individual differences* of learners moderate the learning process. The findings about expert knowledge suggests that *incorporating new information into existing knowledge structures* in a meaningful manner is a key aspect of learning. We discuss these concepts in more detail in the following sections.

2.3 Meaningful Learning as Sensemaking

In this section, we discuss how meaningful learning can be further qualified using the concepts of sensemaking. **Sensemaking**[3] is a process that occurs when learners *connect* their *previously developed* knowledge, ideas, abilities, and experiences together to address the uncertainty presented by a newly introduced phenomenon, problem, or piece of information (Next Generation Science Standards, 2021). A significant portion of learning is sensemaking, especially those which use recorded information or systematic discovery to learn concepts, ideas, theories, and facts in a domain (such as science or history) (P. Zhang & Soergel, 2014). The phrase "figure something out" is often synonymous with sensemaking. Sensemaking is generally about actively trying to figure out the way the world works, and/or exploring how to create or alter things to achieve desired goals (Next Generation Science Standards, 2021). (Dervin & Naumer, 2010) distinguish work on sensemaking in four fields: "Human Computer Interaction (HCI) (Russell's sensemaking); Cognitive Systems Engineering (Klein's sensemaking); Organizational Communication (Weick's sensemaking; Kurtz and Snowden's sense-making); and Library and Information Science (Dervin's sense-making)".

Many theories of learning and sensemaking revolve around the concept of fitting new information into an existing or adapted knowledge structure (P. Zhang & Soergel, 2014). The central idea is that knowledge is stored in human memory as *structures* or *schemas*, which comprise interconnected concepts and relationships. When new information is encountered or acquired, the learner or sensemaker needs to actively construct a revised or entirely new knowledge structure. Examples of some such theories include: the *assimilation theory (theory of meaningful learning)* (Ausubel et al., 1968; Ausubel, 2012; Novak, 2002; Novak, 2010); the *schema theory* (Rumelhart & Norman, 1981; Rumelhart & Ortony, 1977); and the *generative learning theory* (Grabowski, 1996; Wittrock, 1989); all of which have their foundations in the Piagetian concepts of *assimilation* and *accommodation* (Piaget, 1936).

[3]"Brenda Dervin, one of the originators of the sense-making methodology, prefers the spelling with a hyphen, while the community in computer science and more technical people in information science (e.g., SIGCHI) use sensemaking without a hyphen" (P. Zhang & Soergel, 2014).

Assimilation means addition of new information into an existing knowledge structure. A "synonym" (Vakkari, 2016) for assimilation is **accretion**, which is the gradual addition of factual information to an existing knowledge structure, without structural changes. Accretion does not change concepts and their relations in the structure, but may populate a concept with new instances or facts. **Accommodation** means modifying or changing existing knowledge structures, by adding or removing concepts and their connections in the knowledge structure. Accommodation is subdivided into *tuning* / *weak-revision*, and *restructuring*, based on the degree of structural changes (P. Zhang & Soergel, 2014). **Tuning** or **weak revision** does not include replacing concepts or connections between concepts in the structure, but tuning of the scope and meaning of concepts and their connections. This may include, for example, generalizing or specifying a concept. **Restructuring** means radically changing and replacing concepts and their connections in the existing knowledge structure, or creating of new structures. Such radical changes often take place when prior knowledge conflicts with new information. New structures are constructed either to reinterpret old information or to account for new information (Vakkari, 2016; P. Zhang & Soergel, 2014). A comparison of these types of conceptual changes can be found in (P. Zhang & Soergel, 2014 Table 3).

2.3.1 Concept Maps to enhance Sensemaking

As we saw in the previous section, deep learning / meaningful learning / sensemaking is a process in which new information is connected to a relevant area of a learner's existing knowledge structure. However, the *learner must choose* to do this, and must actively seek a way to integrate the new information with existing relevant information in their cognitive structure (Ausubel et al., 1968; Novak, 2010). Learning facilitators (e.g., teachers) can encourage this choice by using the concept mapping technique.

A **concept-map** is a two-dimensional, hierarchical node-link diagram (a *graph* in Computer Science parlance) that depicts the structure of knowledge within a discipline, as viewed by a student, an instructor, or an expert in a field or sub-field. The map is composed of concept labels, each enclosed in a box (graph *nodes*); a series of labelled linking lines (*labelled edges*);

and an inclusive, general-to-specific organization (Halttunen & Jarvelin, 2005). Concept-maps assess how well students see the ''big picture", and where there are knowledge-gaps and misconceptions. A *mind map* is a diagram similar to a concept map, comprising nodes and links between nodes. However, mind maps emerge from a single centre, and have a more hierarchical, tree like structure. Concept maps are more free-form, allowing multiple hubs and clusters. Also, mind-maps have unlabelled links, and are subjective to the creator. There are no "correct" relationships between nodes in a mind map. Figure shows the key features of a concept map, with the help of a concept map.

Concept maps are therefore, arguably the most suited mechanism to represent the cognitive knowledge structures, connections, and patterns in a learner's mind. Conventional tests, such as multiple choice questions, are best at assessing students' recall of facts and guessing skills. Their format treats information as distinct and separate items, rather than interconnected pieces of a bigger picture. Concept maps on the other hand, encourage learners to identify and make connections between concepts that they know, and concepts that are new to them. Concept maps have been used for over 50 years to provide a useful and visually appealing way of illustrating and assessing learners' conceptual knowledge (Egusa et al., 2010, 2014a, 2014b, 2017; Halttunen & Jarvelin, 2005; Novak, 2010; Novak & Gowin, 1984).

Analysis of concept maps can reveal interesting patterns of learning and thinking. Some of these measures that have been used by (Halttunen & Jarvelin, 2005) are: addition, deletion, and differences in top-level concept-nodes; depths of hierarchy; and number of concepts that were ignored or changed fundamentally. In this regard, (Novak & Gowin, 1984) have presented well-established scoring schemes to evaluate concept-maps: 1 point is awarded for each correct relationship (i.e. concept–concept linkage); 5 points for each valid level of hierarchy; 10 points for each valid and significant cross-link; and 1 point for each example.

Having discussed how deep learning / meaningful learning / sensemaking involves creation of knowledge structures in the learner's mind, and suitably adding new pieces of information in the knowledge structure, we now discuss how these processes are influenced in the 21st century with the presence of new media, digital technologies, and information retrieval systems.

2.4 'New' Learning as Online Information Searching

Digital media technologies and e-learning 'ecologies' can enable new forms and models of learning, that are fundamentally different from the traditional classroom practices of didactic pedagogy (Cope & Kalantzis, 2017). Some key concepts associated with these forms of 'new learning' are described below. These concepts from the Educational Sciences domain tie back strongly to the issues, challenges, and research agenda being investigated by researchers in the Search as Learning and Information Retrieval domain (Section 1.1.

2.4.1 Active Knowledge Making

The Internet and new forms of media provide us the opportunity to create learning environments where learners are no longer mainly *consumers* of knowledge, but also *modifiers*, *producers*, and *exchangers* of knowledge. In **active knowledge making**, learners can, and often need to, find information on their own using online resources. They are not restricted to the textbook alone. The Internet is often a definitive resource for information on any given topic. A learner can search the web (to learn) at any time, from anywhere, on any web-enabled device.

As knowledge producers, learners search and analyze multiple sources with differing and contradictory perspectives, and develop their own observations and conclusions. In this process, they become researchers themselves and learn to collaborate with peers in knowledge production. Collaboration gives learners the opportunity to work with others as coauthors of knowledge, peer reviewers, and discussants to completed works. Because learners bring their own views, outlooks, and experiences, the knowledge artefact they create is often uniquely voiced instead of a templated "correct" response (Amina, 2017).

> *Learners become **active knowledge producers** (for instance, project-based learning, using multiple knowledge sources, and research based knowledge making), and not merely knowledge consumers (as exemplified in the 'transmission' pedagogies of traditional textbook learning or e-learning focused on video or e-textbook delivery). Active knowledge making practices underpin contemporary emphases on innovation, creativity and problem solving, which are quintessential 'knowledge economy' and 'knowledge society' attributes.* — Cope & Kalantzis (2017)

2.4.2 Artefacts for Learning Assessment

Traditionally, the focus of learning outcomes has been long term memory. Students and learners were expected to remember a collection of facts, definitions, proofs, equations, and other associated details. For a significant amount of modern knowledge-work today, **memory is actually less important**. Information is so readily accessible now that it is no longer necessary to remember the information. Because of the technological phenomenon, the mass of information is available ubiquitously [4] to a learner (or a knowledge worker), in every moment of learning. Empirical details such as facts, definitions, proofs, or equations do not need to be remembered today, because they can always be looked up again (Amina, 2017; Cope & Kalantzis, 2017).

This creates an interesting shift in the focus of learning and knowledge work today: *"if we are not going to measure and value long-term memory in education, what are we going to assess?"* Cope & Kalantzis (2017) suggest that **we assess the knowledge artefacts** that learners produce. In active knowledge making, the final work [5] can be proof of the learning outcome and represent a learner's ability to use the resources that are available (Amina, 2017). **Measure of learning can be measure of information quality and information use in artefacts.** This shows a shift in pedagogy and assessment and an increase in personalization and individualization of learning (Pea & Jacks, 2014). Memorizing the information on a topic is less important, compared to the writing, synthesizing, analyzing, and **sensemaking** of the available information that has been referenced in the work. This shifts the focus of assessment to the quality of the artefacts and the processes of their construction. Moreover, as technology increases the ability to capture detailed data from formal and informal learning activities, it can give us a new view of how learners progress in acquiring knowledge, skills, and attributes (DiCerbo & Behrens, 2014). Because learning is a continuous, longitudinal process, these advanced, technologically enhanced assessments are more useful in understanding the learning process and knowledge development (Amina, 2017).

[4] as long as there is internet connection

[5] be it a project report, poster, presentation, video, software,research paper, website, etc.

Assessing open-ended artefacts does come with its challenges and limitations. First, assessing and grading artefacts requires the development of detailed qualitative coding guides (M. J. Wilson & Wilson, 2013). This process involves defining grading criteria and measuring inter-coder agreement to ensure that the coding guide is reliable. Prior studies have scored summaries along dimensions such as the inclusion of facts, relationships between facts, and evaluative statements (Lei et al., 2015; Roy et al., 2021; M. J. Wilson & Wilson, 2013). Second, the quality of responses may be difficult to compare across learners. Since this type of assessment imposes very few constraints on the learners' responses, it may cause some learners to *satisfice*, and not convey everything that was learned. Additionally, writing skills are likely to vary across learners, and some may not be able to effectively articulate everything that was learnt.

2.4.3 'Information Search and Evaluation' as and for Learning

Learning today is more about **navigation, discernment, induction, and synthesis**, and less about memory and deduction (Cope & Kalantzis, 2013). However, knowing the source, finding the source, and using the information critically is important to learn and know now more than ever before (Amina, 2017). Learners must know the social sources of knowledge and understand and correctly use quotations, paraphrases, remixes, links, citations, and the like in the works that they develop. Searching and sourcing from the web entails a process of developing and completing a work that inevitably makes learners **knowledge producers**, as long as they can navigate and critically discern the value of multiple sources. This is a skill that must be learned, as many sources of information are not valid, reliable, or authentic (McGrew et al., 2018; Wineburg & McGrew, 2016). Understanding the different sources and identifying the more reliable ones are essential for effective teaching and learning (McGrew et al., 2017; McGrew, 2021). This is a critical aspect because the inability to cite properly or to use reliable resources provides learners with misconstrued information and ideas (Amina, 2017; Breakstone et al., 2021; McGrew et al., 2017).

The Stanford History Education Group (SHEG) conceptualised the **Civic Online Reasoning** (COR) curriculum [6] to enable students to effectively search for and evaluate online information (Breakstone et al., 2018; Breakstone et al., 2021; McGrew, 2020). The curriculum centres on asking three questions of any digital content: *(i)* who is behind a piece of information? *(ii)* what is the evidence for a claim? *(iii)* what do other sources say? The curriculum has lessons and assessments for information evaluation skills such as lateral reading (Wineburg & McGrew, 2017), identifying news versus opinions, checking domain names, identifying sponsored content, evaluating evidence, and practising click restraint (McGrew & Glass, 2021). The lessons were developed and piloted by the Stanford History Education Group (McGrew et al., 2018; McGrew, 2020; McGrew & Glass, 2021). Taken together, these strategies will allow academics and students to better evaluate digital content, from the perspectives of professional fact checkers.

The purview of the *Civic Online Reasoning* curriculum is more targeted than the expansive fields of media and digital literacy [7], (which can embrace topics ranging from cyberbullying to identity theft). Civic Online Reasoning focuses squarely on how to sort fact from fiction online, a prerequisite for responsible civic engagement in the twenty-first century (Breakstone et al., 2021; Kahne et al., 2012; Mihailidis & Thevenin, 2013).

2.5 Promoting Better Learning

> *It is not the technology that makes a difference; it is the pedagogy.*
>
> — Cope & Kalantzis (2017)

Having discussed how meaningful learning takes place, and how it is influenced by the presence of digital media and the mass of information on the Internet, let us now look deeper into the learners as persons themselves. In this section, we discuss how different cognitive and

[6] `https://cor.stanford.edu`

[7] *"Digital literacy describes a holistic approach to cultivating skills that allow people to participate meaningfully in online communities, interpret the changing digital landscape, understand the relationships between systemic -isms and information, and unlock the power of digital tools for good. This includes media literacy. Terms like critical media literacy, media literacy, news literacy, and more are not necessarily interchangeable."* – Collins (2021)

metacognitive practices and aspects of learners can promote better learning. These phenomena have important implications for any digital systems that aim to foster learning.

2.5.1 Externalization and Articulation

The learning sciences have discovered that when learners externalize and articulate their developing knowledge, they learn more effectively (National Research Council, 2000). Best learning takes place when learners articulate their unformed and still developing understanding, and continue to articulate it throughout the process of learning. This phenomenon was first studied in the 1920s by Russian psychologist Lev Vygotsky. Articulating and learning go hand in hand, in a mutually reinforcing feedback loop. Often learners do not actually learn something until they start to articulate it. While thinking out loud, they learn more rapidly and deeply than while studying quietly (Sawyer, 2005). The learning sciences community is actively researching how to support students in their ongoing process of articulation, and which forms of articulation are the most beneficial to learning. Articulation is more effective if it is scaffolded – channelled so that certain kinds of knowledge are articulated, and in a certain form that is most likely to result in useful reflection (Sawyer, 2005). Students need help in articulating their developing understandings, as they do not yet know how to think about thinking, or talk about thinking; their knowledge state is *anomalous* (Belkin et al., 1982).

2.5.2 Metacognition and Reflection

One of the reasons that articulation is so helpful to learning is that it promotes *reflection* or *metacognition*. **Metacognition**, commonly referred to as thinking about thinking, involves thinking at a higher level of abstraction, which in turn improves thinking and learning (Blanken-Webb, 2017). It is “the process of reflecting on and directing one’s own thinking” (National Research Council, 2000, p. 78), and involves thinking about the process of learning, and thinking about knowledge. This ties forward to the self-regulation that effective learners exhibit (Section 2.5.4). Effective learners are aware of their learning process, and can measure how efficiently they are learning as they study.

<table>
<tr><th colspan="2">Components of Metacognition</th></tr>
<tr><th>Knowledge about Cognition</th><th>Regulation of Cognition</th></tr>
<tr>
<td>Declarative Knowledge:
• knowledge about one's skills, intellectual resources, and abilities as a learner
• the factual knowledge the learner needs before being able to process or use critical thinking related to the topic
• students can obtain declarative knowledge through presentations, demonstrations, discussions

Procedural Knowledge:
• knowledge about <u>how</u> to implement learning procedures (e.g., strategies)
• requires students know the process as well as when to apply process in various situations
• students can obtain procedural knowledge through discovery, cooperative learning, and problem solving

Conditional Knowledge:
• knowledge about <u>when</u> and <u>why</u> to use learning procedures
• the determination under what circumstances specific processes or skills should <u>transfer</u>
• students can obtain conditional knowledge through simulation</td>
<td>Planning:
• planning, goal-setting, and allocating resources prior to learning

Information Management:
• skills and strategy sequences used to process information more efficiently (e.g., organizing, elaborating, summarizing, selective focusing)

Monitoring:
• assessment of one's learning or strategy use

Debugging:
• strategies to correct comprehension and performance errors

Evaluation:
• analysis of performance and strategy effectiveness after a learning episode</td>
</tr>
</table>

Figure 2.3: Operational definitions and features of the metacognition components, adapted from Schraw & Dennison (1994) and Vancouver Island University (2021).

The literature on metacognition broadly identifies two fundamental components of metacognition: knowledge about cognition, and regulation of cognition. **Knowledge about cognition** includes three subprocesses that facilitate the *reflective* aspect of metacognition: declarative knowledge (knowledge about self and about strategies), procedural knowledge (knowledge about how to use strategies), and conditional knowledge (knowledge about when and why to use strategies). **Regulation of cognition** include a number of subprocesses that facilitate the *control* aspect of learning. Five component skills of regulation have been discussed extensively in the literature, including planning, information management strategies, comprehension monitoring, debugging strategies, and evaluation. The operational definitions of these components are described in Figure 2.3

Schraw & Dennison (1994) developed the **Metacognitive Awareness Inventory** (MAI) survey and a scoring guide to measure these self-reported components and subprocesses of

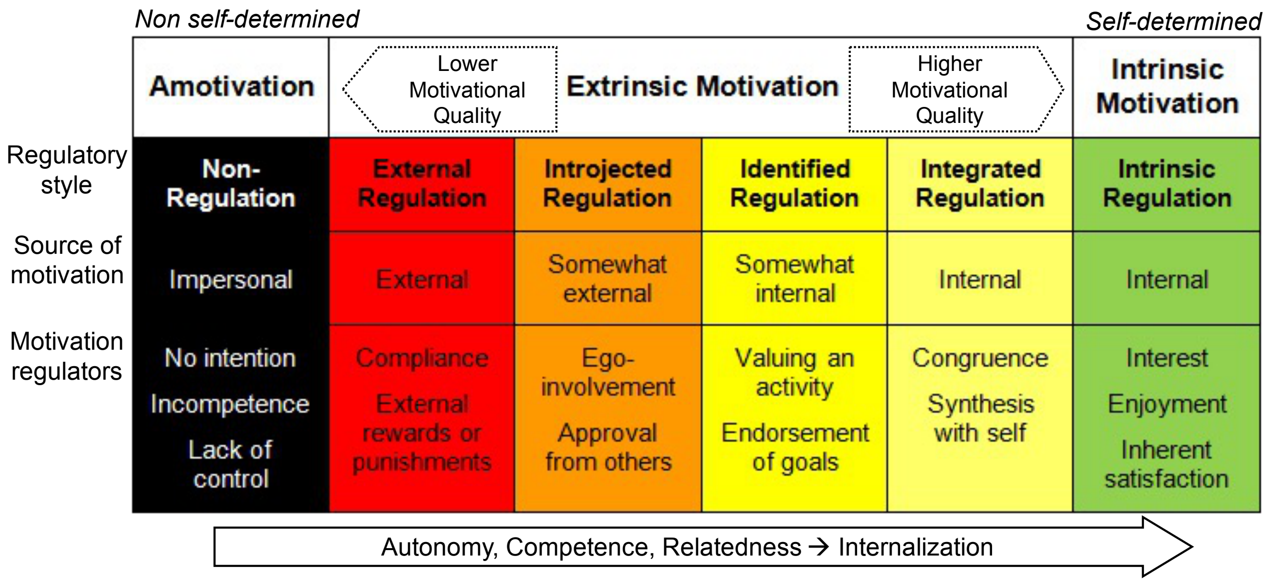


Figure 2.4: The motivation and self-determination continuum, as proposed by the Self-Determination Theory (SDT). Figure adapted from Ryan & Deci (2000a), Ryan & Deci (2000b), and Guyan (2013).

metacognition. The original survey consists of 52 true/false questions (Appendix B.5), such as "*I consider several alternatives to a problem before I answer*", "*I understand my intellectual strengths and weaknesses*", "*I have control over how well I learn", and "I change strategies when I fail to understand*". The instrument has been widely used in research, and has its reliability and validity measures available. Later, Terlecki & McMahon (2018) proposed a revised version of the MAI, using five-point Likert-scales, ranging from "*I never do this*" to "*I do this always*". They argue that when measuring change in metacognition over time, the Likert-scale based 'how often' questions are more effective than dichotomous 'Yes/No' questions (Terlecki, 2020; Terlecki & McMahon, 2018).

2.5.3 Motivation

Motivation is the process that initiates, guides, and maintains goal-oriented behaviours (Cherry, 2020). The **Self-Determination Theory** (SDT) represents a broad framework for the study of human motivation and personality (Ryan & Deci, 2017). SDT differentiates the types of motivation based on the reasons that give rise to behaviour: intrinsic motivation and extrinsic motivation. **Intrinsic motivation** is engaging in a task or behaviour for the rewards *inside* the task or behaviour, such the pleasure, enjoyment and satisfaction that the behaviour provides. It is a stable form of motivation. **Extrinsic motivation** is engaging in a task or behaviour for the rewards *outside* the task or behaviour, such as receiving rewards, avoidance of punishment, gaining social approval, or achievement of a valued result. Extrinsic motivation is on a continuum from less stable to more stable, as illustrated in Figure 2.4. Extrinsic motivation does not last unless the rewards and punishments are explicitly visible (Deci & Ryan, 2013; Ryan & Deci, 2000b; Tahamtan, 2019).

Ryan (1982) proposed the **Intrinsic Motivation Inventory** (IMI) (Appendix B.3), a multidimensional questionnaire intended to assess participants' subjective experience related to a target activity in laboratory experiments. The instrument assesses participants' interest/enjoyment, perceived competence, effort, value/usefulness, felt pressure and tension, and perceived choice while performing a given activity, yielding six subscale scores. The *interest/enjoyment* subscale is considered the most indicative self-report measure of intrinsic motivation. The *perceived choice* and *perceived competence* concepts are theorized to be positive predictors of both self-report and behavioral measures of intrinsic motivation. The *pressure/tension* is theorized to be a negative predictor of intrinsic motivation. *Effort* is a separate variable that is relevant to some motivation questions, so it is used if its relevant. The *value/usefulness* subscale is used to measure internalization, with the idea being that people internalize and become self-regulating with respect to activities that they experience as useful or valuable for themselves.

2.5.4 Self-regulation

Self-regulation is the ability to develop, implement, and flexibly maintain planned behaviour in order to achieve one's goals. Self-regulation, and more broadly, self-direction, are critical to being an effective "lifelong" learner. Self-regulation becomes increasingly important at higher levels of education and in professional life, as people take on more complex tasks and greater responsibilities for their own learning. However, these metacognitive skills tend to fall outside the content area of most courses, and therefore, often neglected in instruction (Ambrose et al., 2010, p. 191). Building on the foundational work of Kanfer (1970b); Kanfer (1970a), Miller and

Brown formulated a seven-step model of self-regulation (J. Brown, 1998; W. R. Miller & Brown, 1991). In this model, behavioural self-regulation may falter because of failure or deficits at any of these seven steps: *(i)* receiving relevant information, *(ii)* evaluating the information and comparing it to norms, *(iiii)* triggering change, *(iv)* searching for options, *(v)* formulating a plan, *(vi)* implementing the plan, and *(vii)* assessing the plan's effectiveness (which recycles to steps *(i)* and *(ii)*). Although this model was developed specifically to study addictive behaviours, the self-regulatory processes it describes are meant to be general principles of behavioural self-control. J. M. Brown et al. (1999) developed the **Self-Regulation Questionnaire** (SRQ) (Appendix B.4) to assess these self-regulatory processes through self-report. The items were developed to mark each of the seven sub-processes of the W. R. Miller & Brown (1991) model, forming seven subscales of the SRQ. The 63-item scale elicits responses in the form of 5-point Likert scale, ranging from strongly disagree to strongly agree. Based on clinical and college samples, the authors tentatively recommend a score of 239 and above as high (intact) self-regulation capacity (top quartile), 214-238 as intermediate (moderate) self-regulation capacity (middle quartiles), and 213 and below as low (impaired) self-regulation capacity (bottom quartile).

2.5.4.1 Self-directed and Self-regulated Learning

As we saw in the previous sections, self-regulation, motivation, and metacognition are key concepts that moderate the learning process. These terms are couched in the concepts of self-regulated learning and self-directed learning.

Self-directed learning (SDL) is a "process in which individuals take the initiative, with or without the help from others, in diagnosing their learning needs, formulating goals, identifying human and material resources, choosing and implementing appropriate learning strategies, and evaluating learning outcomes"(Knowles, 1975, p. 18). **Self-regulated learning** (SRL) can be described as the degree to which students are "metacognitively, motivationally, and behaviourally active participants in their own learning process" (Zimmerman, 1989, p. 329).

Often used interchangeably, self-directed learning (SDL) and self-regulated learning (SRL) have some important similarities and differences (Figure 2.5) (Saks & Leijen, 2014). SDL,

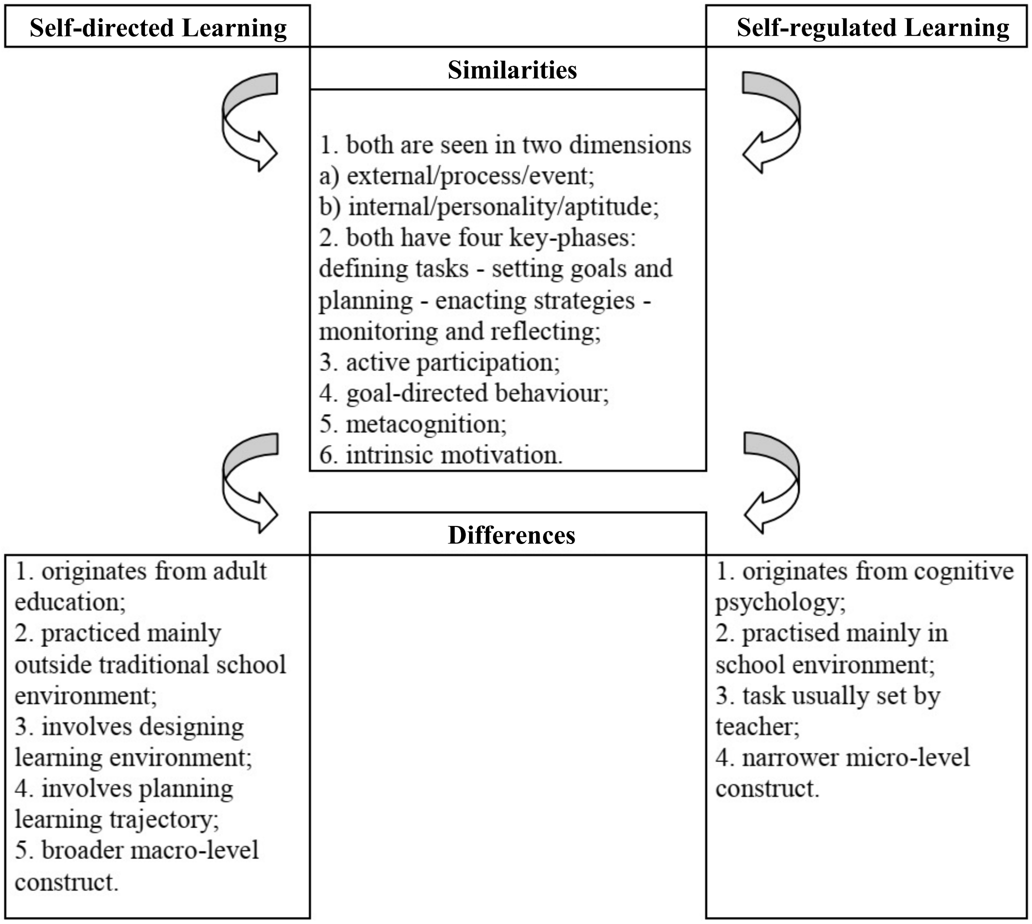


Figure 2.5: Self-directed learning vs. self-regulated learning, as illustrated by Saks & Leijen (2014).

originating from adult education, is a broader, macro-level construct, and is usually practised outside the traditional school environment. The self-directed learner is free to design their own learning environment, and free to plan and set their own learning goals. SRL, on the other hand, is a narrower, micro-level construct, originating from educational and cognitive psychology, and is mostly utilized in the school environment. Learners do not have as much freedom as in SDL. The instructor or facilitator often defines the learning task and the learning goals. Self-directed learning may include self-regulated learning, but the converse is not true (Jossberger et al., 2010; Loyens et al., 2008). In other words, *"a self-directed learner is supposed to self-regulate, but a self-regulated learner may not self-direct"* (Saks & Leijen, 2014). Despite their differences, SDL and SRL share key similarities (Saks & Leijen, 2014). First, both can be seen in two dimensions: *(i) external* to the learner, as a process or series of events, and *(ii) internal* to the learner, arising from the learner's personality, aptitude, and individual differences. Second, both the learning processes have four key phases: *(i)* defining tasks, *(ii)* setting goals and planning, *(iii)* enacting strategies, and *(iv)* monitoring and reflecting. Third, both SDL and SRL require active participation, goal-directed behaviour, metacognition, and intrinsic motivation.

In summary, metacognition is monitoring and controlling what is in the learner's head; self-regulation is monitoring and controlling how the learner interacts with their environment; self-regulated learning is the application of metacognition and self-regulation to learning (Mannion, 2020); and the whole learning process is sustained by motivation, which is desirable to be intrinsic.

2.6 Summary and Implications for this Dissertation

In this first chapter of the background literature review, we discussed *(i)* what is meaningful learning, a.k.a. deep learning, or sensemaking; *(ii)* how meaningful learning updates the learner's cognitive knowledge structure; *(iii)* how the learning process is influenced by digital technologies, mass of information on the Internet, and IR systems; and *(iv)* what principles and practices learners and educators must realize and follow to promote meaningful learning. These findings are from the domains of Educational Sciences, Learning Sciences and Cognitive Sciences. We argue that these are important aspects to be considered when designing future IR or educational information systems that aim to combine and improve the searching and learning experience.

Guided by these findings, we made some important decision choices for the longitudinal study conducted in this dissertation. We aimed to situate learners in their context, and incorporate their individual differences using metacognition, motivation, and self-regulation characteristics. Additionally, we aimed to assess learning using artefacts and concept maps. We chose not use traditional tests like question-answers, and multiple choice assignments, since they are often not the preferred choice of knowledge-work output in real world scenarios.

In the next chapter, we look at relevant literature from the Information Sciences and Interactive Information Retrieval disciplines.

3
Background: Information Searching

This second chapter on background literature discusses relevant concepts from the disciplines of Information Sciences, and more specifically Interaction Information Retrieval. First, we introduce some terminology around information behaviour, information need, and information relevance. Then we discuss relevant findings various empirical studies, from the lens of three-stage interactions in the information search process. Then we discuss some overall generic characteristics of information search behaviour, and how they are linked to expertise and working memory. Next we discuss how learning has been assessed in recent search-as-learning studies. We also discuss some limitations of current search systems to foster learning, including the lack of sufficient number of longitudinal studies. In the last section, we state what implications these findings had for shaping the longitudinal study conducted for this dissertation.

3.1 Terminology

Information retrieval (IR) is the process of obtaining *information objects*, that are *relevant* to an *information need*, from a collection of those objects (Wikipedia). **Information objects** are entities that can potentially convey information. They can take many forms, such as documents, webpages, facts, music, spoken words, images, videos, artefacts, and other forms of

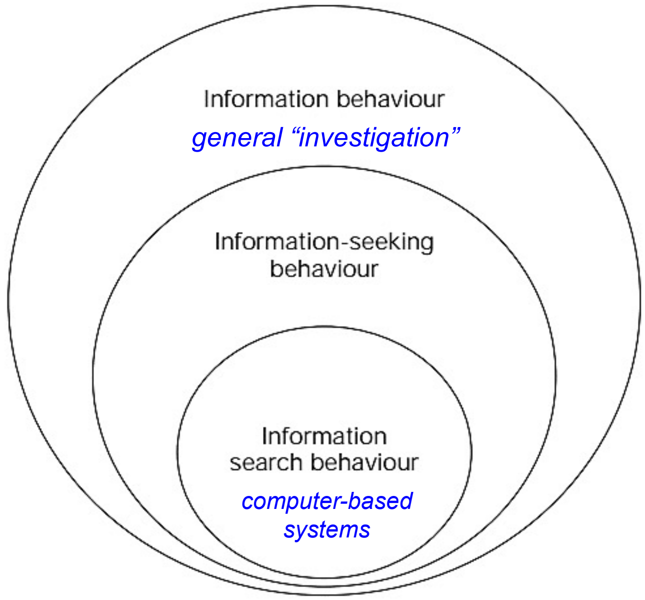


Figure 3.1: Nested model of information behaviour by T. D. Wilson (1999).

human expression. Areas where information retrieval techniques are employed include search engines, such as web search, social search, and desktop search; media search, as in image, music, video; digital libraries and recommender systems, as well as domain specific applications like geographical information systems, e-Commerce websites, legal information search, and others.

Multiple perspectives exist around how users interact with information, and IR systems. In the **Search Engine application view**, the interactions are restricted to the search engine interface. In the **Human-computer interaction** (HCI) view, interactions are between a person and a system; but the system can go *beyond* supporting only retrieval, to supporting more complex tasks. In the **cognitive view of IR**, which is the broadest, the interactions for obtaining information can be between a person and a system, as well as between people, for retrieval of information.

People's behaviour around information can be modelled as a nested Venn diagram as proposed by T. D. Wilson (1999) (Figure 3.1). **Information behaviour** is the more general field of investigation. **Information-seeking behaviour** can be seen as a sub-set of the field, particularly concerned with the variety of methods people employ to discover, and gain access to information objects. **Information search behaviour** is yet a sub-set of information-seeking, concerned with the interactions between the user and computer-based information systems. In this dissertation, we focus on information search rather than the other two higher hierarchical concepts. This is because online IR systems, such as search engines or digital libraries, have become the primary source for people to obtain information in modern times, and web search is becoming ever more pervasive and ubiquitous in our day-to-day lives.

The field of **interactive information retrieval** (IIR) posits that IR systems should operate in the way that good libraries do. Good libraries provide both the information a visitor needs, as well as a *partner* in the learning process — the information professional — to navigate that information, make sense of it, preserve it, and turn it into knowledge. As early as in 1980, Bertram Brookes stated that searchers acquire new knowledge in the information seeking process (Brookes, 1980). Fifteen years later, Gary Marchionini described information seeking, as *"a process, in which humans purposefully engage in order to change their state of knowledge"* (Marchionini, 1995). So we have known for quite a while that search is driven by the higher-level human need to gain knowledge. Information Retrieval is thus a means to an end, and not the end in itself. Thus, the ideal IR system should not only help users to locate information, but also help them to **bridge the gap between information and knowledge**.

This brings us to the concept of information need. **Information Need** is the desire to locate and obtain information to satisfy a conscious or unconscious human need. Most search systems of today assume that the search query is an accurate representation of a user's information need. However, Belkin et al. (1982) observed that in many cases, users of search systems are unable to precisely formulate what they need. They miss some vital knowledge to formulate their queries. As humans, we have difficulty in asking questions about what we do not know. Belkin called this phenomenon as **Anomalous State of Knowledge**, or ASK. Later, Huang & Soergel (2013) identified an exhaustive set of criteria that should be considered in order to ideally represent a user's information need. These criteria for information need are highly dependent on the user context: user attributes, tasks or goals, as well as the situation the user is embedded in. This brings us to another closely related concept: information relevance.

Relevance is a fundamental concept of Information Science and Information Retrieval, and perhaps the most celebrated work in this area has been done by Tefko Saracevic (Saracevic, 1975, 2007a, 2007b, 2016). Webster dictionary define relevance as "a relation to the matter at hand". In most circumstances, relevance is a "y'know" notion. People apply it effortlessly, without anybody having to define for them what "relevance" is. This creates one of the most fascinating challenges in the information field: humans understand relevance intuitively, while it is an open research problem to represent relevance effectively for use by algorithmic systems. The situation becomes more interesting because relevance always depends on context, and the context is ever dynamic, as the matter at hand changes.

3.2 Three-stage Interactions with Online Search Systems

As we saw in the previous section, information search behaviour is the (study of) interactions between a user, and digital Information Retrieval (IR) systems. The field of Information Science/Studies has developed multiple models explaining how information search works (T. D. Wilson, 1999). A few of them are presented in Figure 3.2. Across many of these models, we observe that most major Information Retrieval (IR) systems have three fundamental ways of letting users interact with the system, and the underlying information: *(1)* an interface for entering search **queries**; *(2)* an interface for viewing and evaluating a **list** of retrieved information-objects, or search results; *(3)* an interface for viewing and evaluating **individual information-objects**. For instance, Marchionini (1995)'s ISP model hints at these three interfaces in the fourth, sixth and seventh stages, namely "formulate query", "examine results", and "extract info". Spink (1997)'s model of the IR interaction process consists of sequential steps or cycles, and each cycle comprises one or more interactive feedback occurrences of user input (query), IR system output (list), and user interpretation and judgement (of individual information-objects). Consequently, findings from the large body of empirical research in interactive IR (especially those with web based search systems) can be grouped around these thee stages of interactions with search systems:

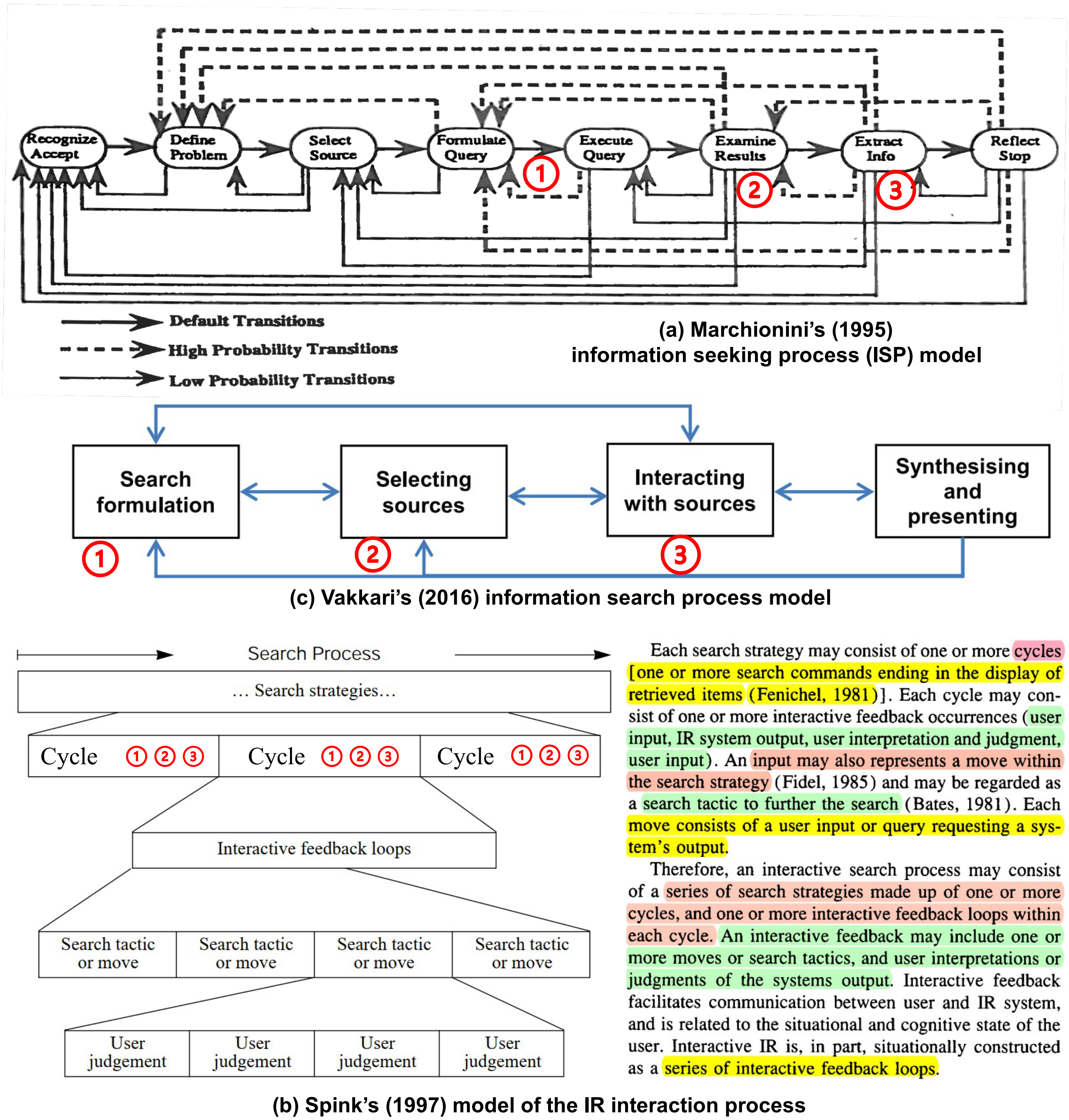


Figure 3.2: Models of information search process, with our coloured annotations identifying the three stages: (1) query formulation, (2) list-item selection, and (3) item examination.

1. *Stage 1:* search query (re)formulation
2. *Stage 2:* list-item selection: search results evaluation (aka source selection)
3. *Stage 3:* item examination: content page evaluation (aka interacting with sources)

The discussions in the following subsections are based around these three stages of interactions. The empirical studies discussed below generally follow some common principles of user studies in Interactive IR (IIR) (Borlund, 2013; Kelly, 2009): participants are presented with a search task or search topic, and then they are asked to search the internet (or a simulation of the open web) for information. During the search, the various interactions (queries, clicks, webpages opened etc.) are recorded, and these are analysed and correlated with other sources of data to answer research questions.

3.2.1 Stage 1: Query (Re)formulation

How do users behave when submitting search queries (to an IR system)?

Query formulation is the process of composing a search query that describes the information need of a searcher. **Query reformulation** refers to the act of either modifying a previous query, or creating a new query. Query reformulation typically occurs due to a searcher's improved understanding of how to better translate their information need into a search query. The relationship between two successively issued queries have been classified in a number of ways. These classifications are called *Query Reformulation Types*, or QRTs. Amongst many other, Boldi et al. (2009) used cognitive aspects of the searchers issuing the query to propose a taxonomy of QRTs, while C. Liu et al. (2010) proposed a similar taxonomy focusing more on the linguistic properties of the two successive queries. These are compared and contrasted in Figure 3.3.

Task-type, task-topic, task-goal, and domain-expertise were found to influence query reformulation patterns of searchers (Eickhoff et al., 2015; Jiang et al., 2014; Mao et al., 2018). At first glance, a significant portion of the query reformulation terms ($\sim 86\%$) seemed to be coming from the task-description itself (Jiang et al., 2014; Mao et al., 2018). This was characterized by

Boldi et al. (2009)	Liu et al. (2010)
Used cognitive aspects of searchers issuing the query:	Used linguistic properties of two successive queries:
Generalization: user wants broader information than was obtained from the current search	**Generalization:** successive queries contain at least one term in common; second query contains fewer terms than first query
Specialization: user narrows down the current search	**Specialization:** successive queries contain at least one term in common; second query contains more terms than first query
Mission Change: user changes the search topic to an entirely different one	**Word Substitution:** successive queries contain at least one term in common; second query has same length as first query, but contains some terms that are not in first query
Parallel Move: user modifies the current query to change the search aspect with the same context Learners are the also knowledge producers, and discerning knowledge discoverers / navigators	**Repeat:** successive queries contain exactly the same terms, but the format or ordering of these terms may be different
Error Correction: user's search intent does not change in the period before and after reformulation; examples are correcting a misspelled term and/or performing a query paraphrase	**New:** successive queries do not contain any common terms

Generalization and *Specialization* are identical in both taxonomies. *Parallel Move* can contain *Word Substitution*, *Repeat*, or *New* QRTs. *Mission Change* will generally have *New* QRTs. *Error Correction* will possibly have *Repeat* or *Word Substitution*.

Figure 3.3: Comparison of Query Reformulation Types (QRTs) proposed by Boldi et al. (2009) and C. Liu et al. (2010).

significantly more fixations on the task-description, rather than other SERP elements. Jiang et al. (2014) and Mao et al. (2018) investigated this phenomenon further. Jiang et al. (2014) controlled for the task-type and task-goal, using the faceted-framework by Li & Belkin (2008). Mao et al. (2018) controlled for the task-topic and the domain-expertise of the searchers.

If search tasks had *factual* goals, searchers relied heavily on the task-description for reformulating their queries (Jiang et al., 2014). For *interpretive* tasks (intellectual tasks with specific goals), users spent more time reading search result surrogates, before reformulating their queries. This was observed by increased eye-fixations (indicative of visual attention) and dwell time on search result snippets (surrogates). For exploratory tasks, searchers fixated the longest on query-autocompletion (QAC) suggestions, indicating that they were possibly looking for help and suggestion based on their specific query, as the search-task had non-specific (amorphous) goals.

Searchers also relied on the task-description for reformulating queries, when the search-task was outside their domain of expertise (Mao et al., 2018). For in-domain tasks, they used query terms from their own knowledge, that were not fixated on in visited SERPs and content pages. Eickhoff et al. (2015) reported that a significant share of new query terms came from visited SERPs and content pages, and query reformulation (specialization) often did not literally re-use previously encountered terms, but highly related ones [1] instead. These observations can possibly be explained by Mao et al. (2018)'s findings: when exploring a new domain, the searcher may accumulate vocabulary and learn how to query during the search; when performing in-domain search-tasks, the searcher may have enough prior knowledge to come up with effective query terms. It was also seen that searchers from medicine domain used more unread query terms for their in-domain search-tasks, compared to politics and environment domains (Mao et al., 2018). This suggested that domain knowledge and expertise is more important for formulating good search queries in highly technical disciplines (e.g., medicine), compared to less technical domains (e.g., politics).

Query Auto Completion (QAC) is a technological feature that suggests possible queries to web search users from the moment they start typing a query. It is nearly ubiquitous in modern search systems, and is thought to reduce physical and cognitive effort when formulating a query. QAC suggestions are usually displayed as a list (Figure 3.4(b) and (c)), and users interact in a variety of ways with the list. Hofmann et al. (2014) observed a strong position bias among searchers who examined the QAC list: the top suggestions received the highest visual attention, even when the ordering of the suggestions were randomized. Average fixation time decreased consistently on suggested items from top to bottom. Even when the ranking of suggestions were randomized, time taken to formulate queries did not significantly differ.

Search topics were found to have a large effect on QAC usage (Jiang et al., 2014; Smith et al., 2016). Search was easiest for the topics with the highest QAC usage. Total eye-gaze duration was longest when visual attention was shared between the QAC suggestions and the actual search query input box. Some additional time was probably due to decision making on whether to use a QAC suggestion. Typing was faster when a QAC was not used. However, the IR system's retrieval performance (measured using NDCG@3), was greater when QAC was

[1] measured using Leacock-Chodorow semantic similarity metric (Leacock & Chodorow, 1998)

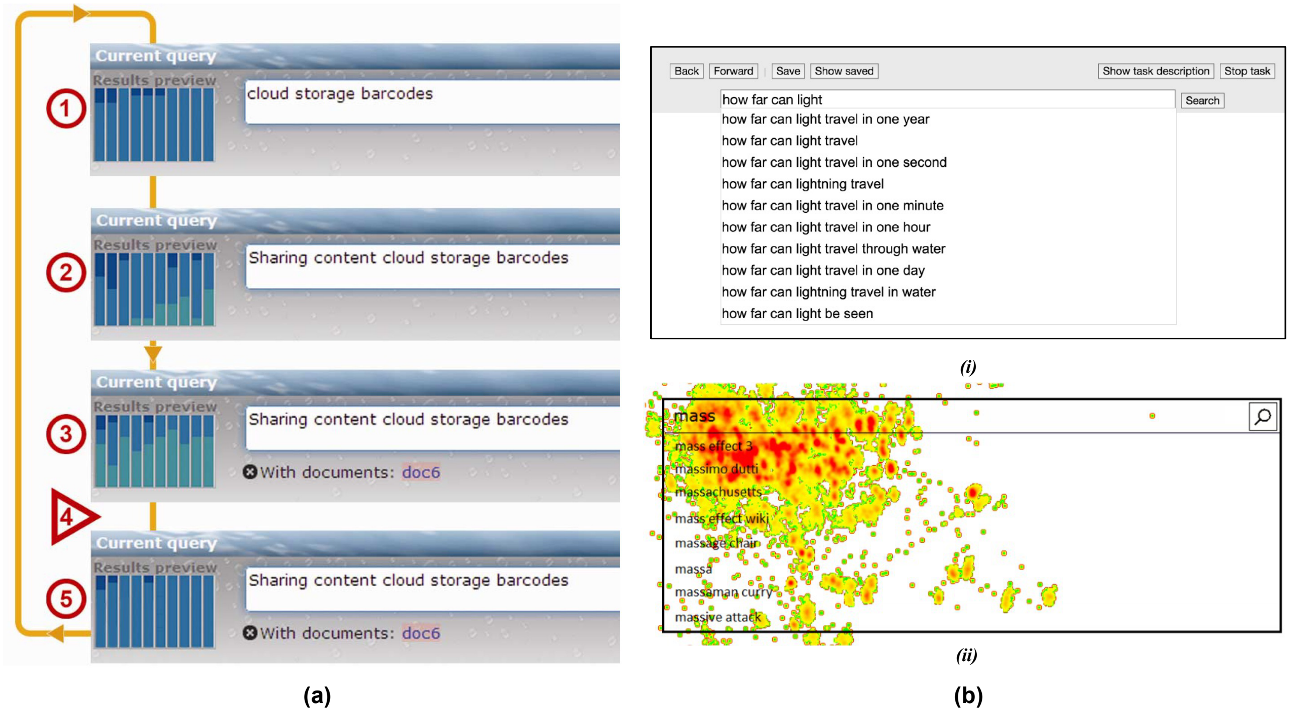


Figure 3.4: Investigating user-interactions with queries: **(a)** Visualizing the distribution of retrieved search results prior to running a query, for helping searchers understand their queries' effectiveness (Qvarfordt et al., 2013). The visualization is a stacked column chart with ten columns. Each column represents ten search results: first column represents results ranked 1-10, second column represents results 11-20, etc. Individual columns have three divisions, indicating the counts of results that: are already seen by the searcher (dark blue, top), will be re-retrieved, but have not been seen by the searcher (medium blue, middle), and will be newly-retrieved (bright teal blue, bottom). The system evaluates the searcher's query continuously as it is being typed, and updates the visualization in real-time. **(b)** Interfaces for examining interactions with query auto-completion (QAC), by ***(i)*** Smith et al. (2016), and ***(ii)*** Hofmann et al. (2014) (overlaid with heatmaps of eye fixations for all participants). This figure is best viewed in colour.

used. So Smith et al. (2016) speculated that the value of using QAC suggestions was realized later in the search session by users, when they saw a reduction in the number of additional queries needed, or an increase in the value of the information found.

Several user behavioural profiles were identified by exploring associations between visual attention from eye-tracking, search interactions from mouse and keyboard activity, and the use of QAC suggestions (Hofmann et al., 2014; Smith et al., 2016). These profiles are described in Figure 3.5. An interesting, yet common-sense observation was that participants' touch-typing ability greatly influenced their interactions with QAC suggestions.

Hofmann et al. (2014) From 331 search 'episodes' (*N* = unmentioned)	**Smith et al. (2016)** From 707 queries submitted in 232 topic sessions (*N* = 29)
A: monitoring: frequent fixations on QAC, and on the top-ranked suggestions in particular (hypothesized that QAC played a role in confirming to the user that they were typing the query correctly) **B: searching:** user actively scanned and engaged with the QAC list from top to bottom; two distinct types of searching were identified: *1)* *seeking spelling support* for difficult words (e.g. 'schwarzenegger') while entering a query they have in mind; 2) not seeking spelling support, but rather *looking for a complete query* that appropriately expressed their information need; **C: ignoring:** non-touch-typists largely ignored the QAC suggestions because they primarily looked at the keyboard while typing, and typically only looked up from the keyboard when they had finished typing.	**A: Fast Boxers:** worked quickly, submitted queries in the query input box, using the enter key almost exclusively, with very little attention to the screen other than to the query input box **B: Touch Typists:** similar to A, worked quickly and mostly using the query input box and the enter key, but not to the exclusion of queries from the QAC suggestions and clicking on them. They looked at the QAC suggestions very rarely, but they did focus on the query input box, as they rarely have a missing fixation. **C: Slow and Methodical Typists:** worked slowly to create many long queries in the query input box. Although they tended to focus a great deal of attention to all parts of the screen, they were likely to focus on the query input box alone for many queries, and rarely focused only on the QAC suggestions list. **D: Agnostic Mousers:** used both the query input box and the QAC suggestions list, but put more time focus on QAC list than the query box. Queries were submitted by clicking the mouse, to the near exclusion of the enter key. A high visual attention to the screen could be explained by the reliance on the mouse and the need to track the mouse cursor visually. **E: Fast and Unfocused Agnostics:** worked quickly to submit a lot of queries using QAC suggestions, but they also typed some queries fully in the query input box. Although a lot of their visual attention was on QAC suggestions list and very little elsewhere, they had a lot of missed fixations, and may have focused on the keyboard. **F: Fast QACers:** also worked quickly, but created very short queries that were likely to be queries from the QAC suggestions. Their attention often focused on the QAC suggestions list only, and rarely on the query input box.

Figure 3.5: Comparison of User behaviour profiles identified around Query Auto-Completion (QAC), from eye-tracking data, by Hofmann et al. (2014) and Smith et al. (2016).

The native language of searchers was found to influence their overall querying and searching behaviour. Ling et al. (2018) explored this space using four variations of a multi-lingual search interface. They observed that participants strongly preferred to issue queries in their first or native language. A second or non-native language was the next preferred choice. Mixing of first and second-languages occurred very rarely. In 80% of the total 300 tasks (25 users $\times$ 4 interfaces $\times$ 3 task-types), participants used a single language for querying. In the rest 20% of the tasks, participants switched languages for querying, with a transition from first language to second language being the most common.

3.2.2 Stage 2: Search Results Evaluation / List-Item Selection

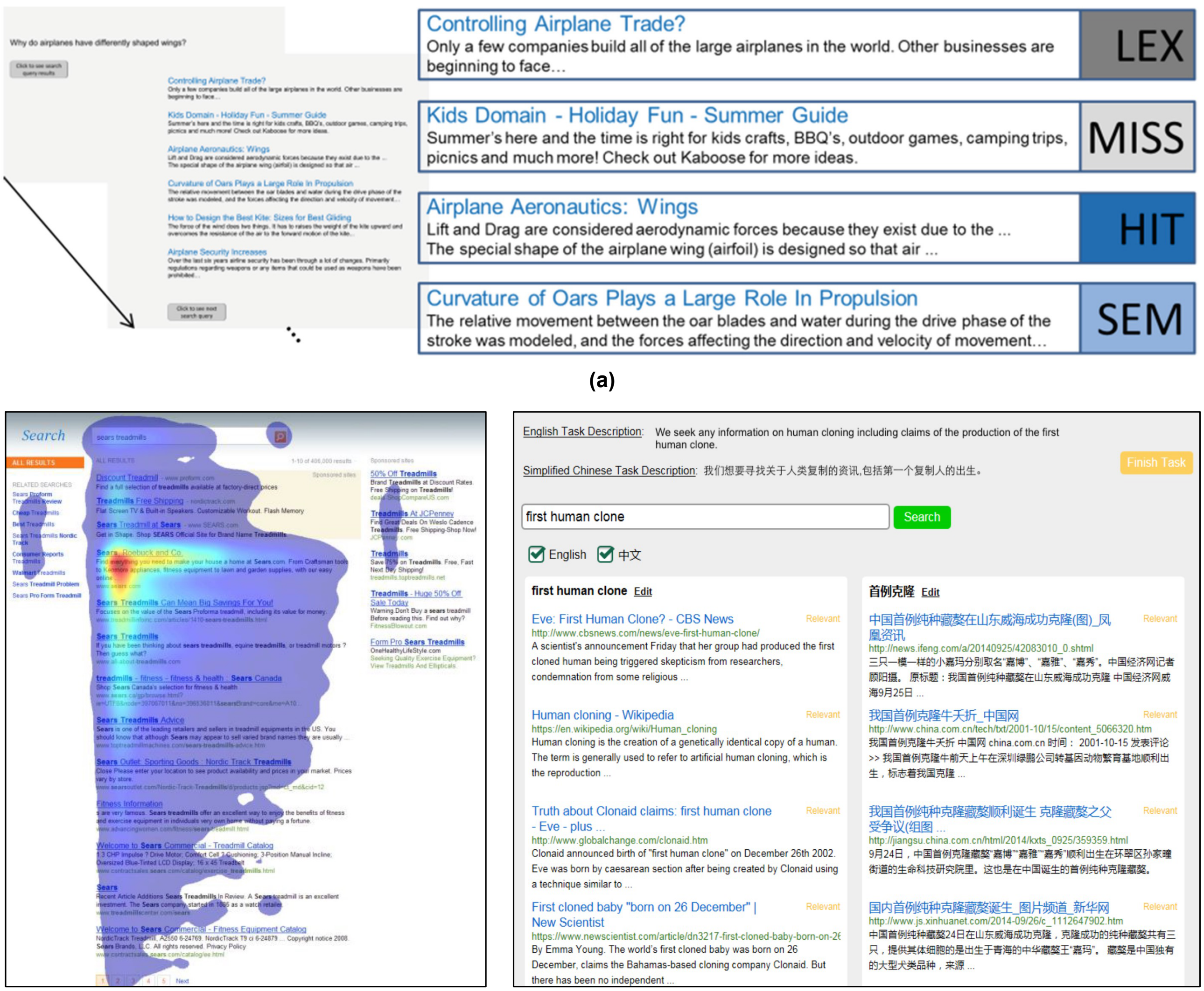
How do users behave when examining a list of information-objects (returned by an IR system)?

After a user submits a query to an IR system, the next action they generally perform is examining and evaluating the list of search results returned by the IR system. In this section, we discuss empirical studies which investigated information-searching behaviour around a list of information-objects, or a representation of information-objects (also called *surrogates*). We identified some common themes in the research questions investigated. The discussion below is grouped along these themes, as relationships between search behaviour and: *(i)* ranking of search results; *(ii)* information shown in search results; *(iii)* individual user characteristics; and *(iv)* relevance judgement and feedback.

3.2.2.1 Ranking of search results

Most search engines display results in a rank ordered list, with the highest *algorithmically* relevant results placed at the top, and others results ordered below. Granka et al. (2004; Lorigo et al., 2008) studied eye-movement behaviour of searchers examining SERPs, and reported observations from three user studies. They saw that in 96% of the queries, participants looked at only the first result page, containing the top 10 results. No participant looked beyond the third result page for a given query. Participants looked primarily at the first few results, with nearly equal attention (dwell time) given to the first and the second results. However, despite equal attention, the first result was clicked 42% of the time, while the second was clicked only 8% of the time. If none of the top three results appeared to be relevant, then users chose not to explore further results, but issued a reformulated query instead. When the ranking of the search results were reversed (i.e. placing less relevant results in the higher ranked positions), participants spent considerably more time scrutinizing and comparing results (more fixations and regressions) before making a decision to click or reformulate.

Some effects of gender were found to influence SERP examination (Lorigo et al., 2008). Females clicked on the second result twice as often, and made more regressions or repeat viewings of already visited abstracts, compared to males. Males were more likely to click



(b) (c)

Figure 3.6: Example interfaces for studying user-interactions with a search-engine results page (SERP): **(a)** a simplified SERP without query input facility, to judge relevance of search results (on a 4-level scale) for pre-determined search queries (in this case 'why do airplanes have differently shaped wings?'), from Scharinger et al. (2016); **(b)** eye-tracking heatmap on an organic SERP from Buscher et al. (2010; Dumais et al., 2010), showing the F-shaped pattern of visual attention; **(c)** a multilingual SERP from Ling et al. (2018). This figure is best viewed in colour.

on lower ranked results, from entries 7 through 10, and also look beyond the first 10 results significantly more often than women. Males were also more linear in their scanning patterns, with less regressions. Pupil dilation did not differ significantly between gender groups.

Effects of task-type and task-goals also influenced SERP examination behaviour. Guan & Cutrell (2007) used Broder (2002)'s taxonomy of navigational vs. informational searches. The authors reported that when users could not find the target results for navigational searches, they either selected the first result, or switched to a new query. However, for informational searches, users rarely issued a new query and were more likely to try out the top-ranked results, even when those results had lower relevance to the task. This illustrated possible strong confidence of searchers in the search engine's relevance ranking, even though searchers clearly saw target results at lower positions. Thus, people were more likely to deprecate their own sense of objective relevance and obeyed the ranking determined by the search engine. Jiang et al. (2014) used Li & Belkin (2008)'s framework of search-tasks, and saw that in tasks having specific goals, searchers fixated more on lower ranked results after some time. On the other hand, for tasks having amorphous goals, there was a wider breadth in viewing the SERP, and less effort spent in viewing the content pages. Fixations tended to decrease as search session progressed, indicating decreased interest and increasing mental effort, which could demonstrate *satisficing* behaviour (Simon, 1956). A comprehensive overview of various behavioural traits associated with task-types and task-goals can be found in (Jiang et al., 2014 Table 8).

3.2.2.2 Information Shown in Search Results (Surrogates)

The amount and quality of different kinds of information shown on SERPs also affected user's information searching behaviour. Cutrell & Guan (2007) saw that as the length of the surrogate information (result snippets) was increased, user's search performance improved for informational tasks, but degraded for navigational tasks (Broder, 2002). Analyzing eye-tracking data, they posited that the difference in performance was due to users paying more attention to the snippet, and less attention to the URL located at the bottom of the search result. This led to performance deterioration in navigational searches. Buscher et al. (2010) studied the effects of the quality of advertisements placed in the SERPs (Figure 3.6(b)). Similar to findings discussed above, a strong position bias of visual attention was found towards the top few organic result entries — the well known F-shaped pattern of visual attention — which was stronger for informational than for navigational tasks. However, a strong bias *against* sponsored links was observed in general. Even for informational tasks, where participants generally had a harder time finding a solution, the ads did not receive any additional attention from the participants. Lorigo et al. (2008) compared the visual attention patterns of searchers using two different search engines: Google, and Yahoo!. Behavioural trends followed similar patterns for both search engines, even though Google was rated as the primary search engine of all but one of the participants. They found slight variations in some eye-tracking measures (reading time of surrogates, time to click results, and query reformulation time), and some self-reported measures (perceived ease of use, perceived satisfaction, and success rate). However, none of these differences were statistically significant.

The novel query-preview interface by Qvarfordt et al. (2013) was discussed in Section 3.2.1 and in Figure 3.4(a). The authors also reported several observations about user behaviour on SERPs. They saw that the presence of the preview visualization enabled participants to look deeper into the results lists. Participants tried to use the preview as a navigation tool, although it was not designed as such. The tool increased the rates at which participants examined documents at middle ranks in query results, and thus helped discover more useful documents in those middle ranks than without the preview widget. The preview tool also helped to increase the diversity of documents found in a search session, which could in turn lead to better performance in terms of recall and precision. Thus, the tool helped searchers overcome the strong position bias towards top-ranked results, as observed by other studies discussed previously.

3.2.2.3 Individual User Characteristics

Individual traits of searchers also influence their pattern of interactions with a SERP, and these patterns can be revealed by analyzing eye-tracking data. For instance, searchers have been classified as *economic* vs. *exhaustive*, based on their style of evaluating SERPs (Aula

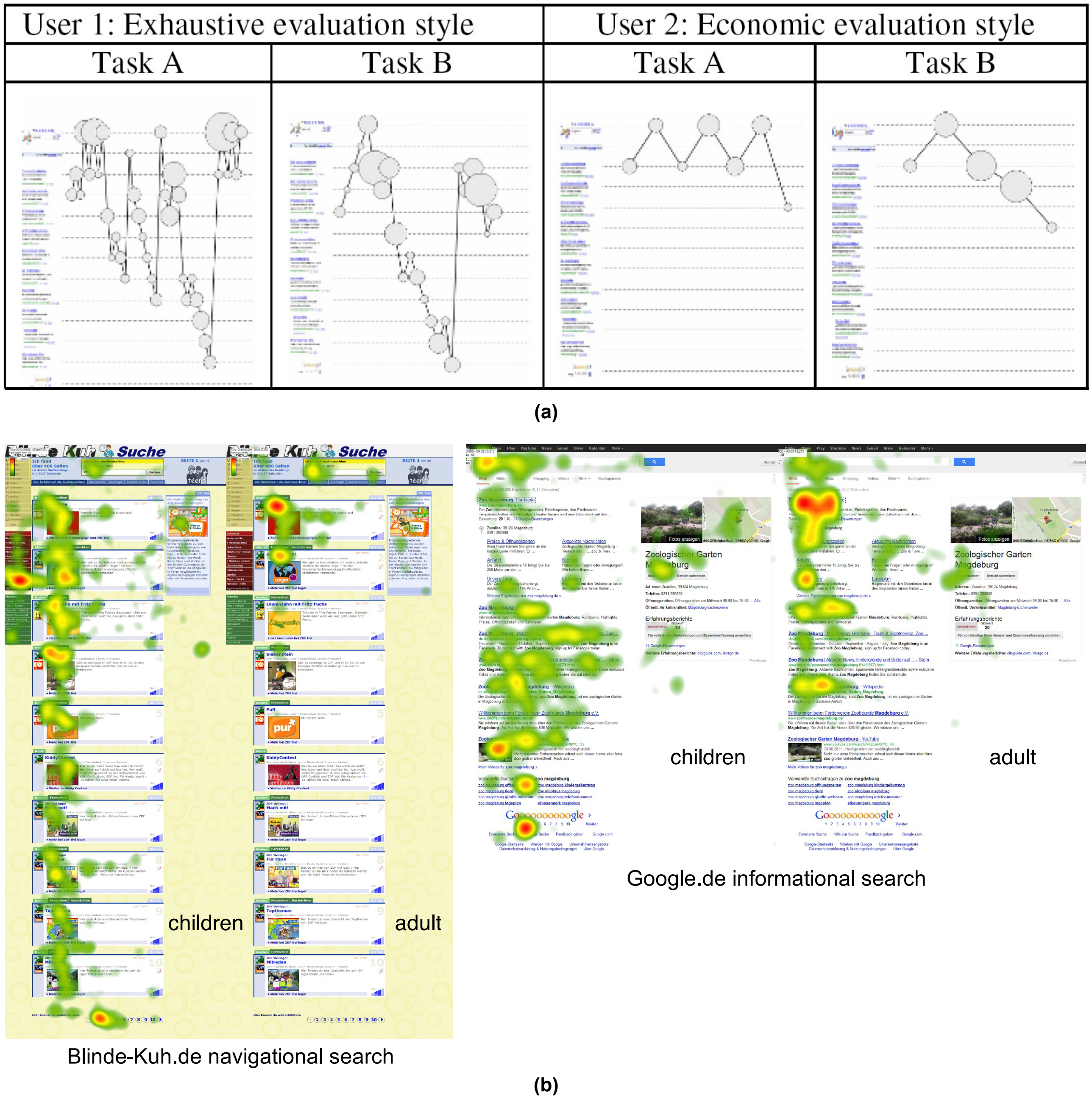


Figure 3.7: Effects of differences in user characteristics on interactions with SERPs: **(a)** exhaustive or *depth-first* user (User 1), vs. economic or *breadth-first* user (User 2), examining mostly irrelevant results in Task A, and mostly relevant results in Task B (both users followed the second link in Task B); vertical axis denotes vertical location on SERP, and horizontal axis denotes temporal ordering of result examination; from Aula et al. (2005); (similar patterns were identified by Bilal & Gwizdka (2016), in the SERP examination behaviour of children) **(b)** children vs. adults examining SERPs from a German search engine for children (left), and Google (right); differently from adults, children exhaustively explored all search results, paid more attention to thumbnails and embedded media, and read less text-only snippets; from Gossen et al. (2014). Similar observations as with children were reported for searchers with dyslexia (Palani et al., 2020). This figure is best viewed in colour.

et al., 2005). *Economic* searchers were found to scan less than half (three) of the displayed results above the fold, before making their first action (query re-formulation, or following a link). *Exhaustive* searchers evaluated more than half of the visible results above the fold, or even scrolled the results page to view all of the results, before performing the first action. Thus, economic searchers demonstrated depth-first search strategy, while exhaustive users favoured the breadth-first approach (Figure 3.7(a)). Dumais et al. (2010) demonstrated the use of unsupervised clustering to re-identify the *economic-exhaustive* user groups, based on differences in total fixation impact [2], scanpaths, task outcomes, and questionnaire data. The *economic* cluster was further broken down by users who looked primarily at results (*economic-results* cluster), and users who viewed both results and ads (*economic-ads* cluster). All three groups spent the highest amount of time on the first three results, with the *exhaustive* group being substantially slower than the other two groups. The *exhaustive* and *economic-results* groups spent the second-highest amount of time on results four through six, while the *economic-eds* group spent this time on the main advertisements. This group spent more than twice as much time on the main ads as the *economic-results* group, and even more time on main ads than the *exhaustive group*. This observation is incongruent to Buscher et al. (2010)'s findings, as they observed a generally strong bias *against* viewing sponsored links. Abualsaud & Smucker (2019) conducted further analysis using these user types, and, in general, reconfirmed the previous findings. They found that the results above the fold, especially, ***the first three search results are special***, more so for economic users. On submitting a 'weak' query, if economic users did not find a correct result within the first three results, they abandoned examination, and reformulated their query.

Age of searchers also influence SERP evaluation behaviour. Gossen et al. (2014) demonstrated differences in SERP evaluation for children and adults (Figure 3.7(b)). When answers were not found within the top search results, the adults reformulated the query starting a new search, while young users exhaustively explored all the ten results, and used the navigation buttons between results pages to continue further examination. Children also paid more

[2] a measure derived from eye fixation durations, proposed by Buscher et al. (2009)

attention to thumbnails and embedded media, and focused less on textual snippets. Children saw the query suggestions at the bottom of the Google SERP (because they navigated to the bottom), while the adults did not. Bilal & Gwizdka (2016; Gwizdka & Bilal, 2017) investigated this phenomenon further, and observed that even within children, age plays a role in SERP evaluation behaviour. Younger children (grade six, age 11) clicked more often on results in lower-ranked positions than older children (grade eight, age 13). Older children's clicking behaviour was based more often on reading result snippets, and not just on the ranked position of a result in a SERP. Whereas, younger children made less deliberate choices in choosing which result to click, and were more exhaustive in the exploration of results. Thus, using Aula et al. (2005)'s classification and Dumais et al. (2010)'s observations, it can be posited that (younger) children start out as *exhaustive* searchers. With increase in age and maturity, older children and adults evolve into *economic* searchers. Interestingly, very similar behaviour patterns as with children (scrolling further down on SERPs, exhaustive exploration, etc.) were also observed recently for searchers with dyslexia (Palani et al., 2020).

Searcher's native language also influenced SERP interaction behaviour (Ling et al., 2018) (Figure 3.6(c)). We discussed in Section 3.2.1 that users strongly preferred issuing queries in a single language, especially their native language. However, while examining SERPs, they marked search results in both their first language and second language to be relevant, to an equal degree. This confirms the usefulness of search result pages that integrate results from multiple languages. However, a clear separation in the language of the search results was strongly preferred, and an 'interleaved' presentation (e.g. odd numbered results in one language and even numbered results in another language) was least preferred.

3.2.2.4 Relevance Judgement

Balatsoukas & Ruthven (2010, 2012) proposed a list of relevance criteria for understanding how searchers evaluate search results, or perform *relevance judgement*. These criteria were developed based on literature reviews and their empirical findings from eye-tracking studies.

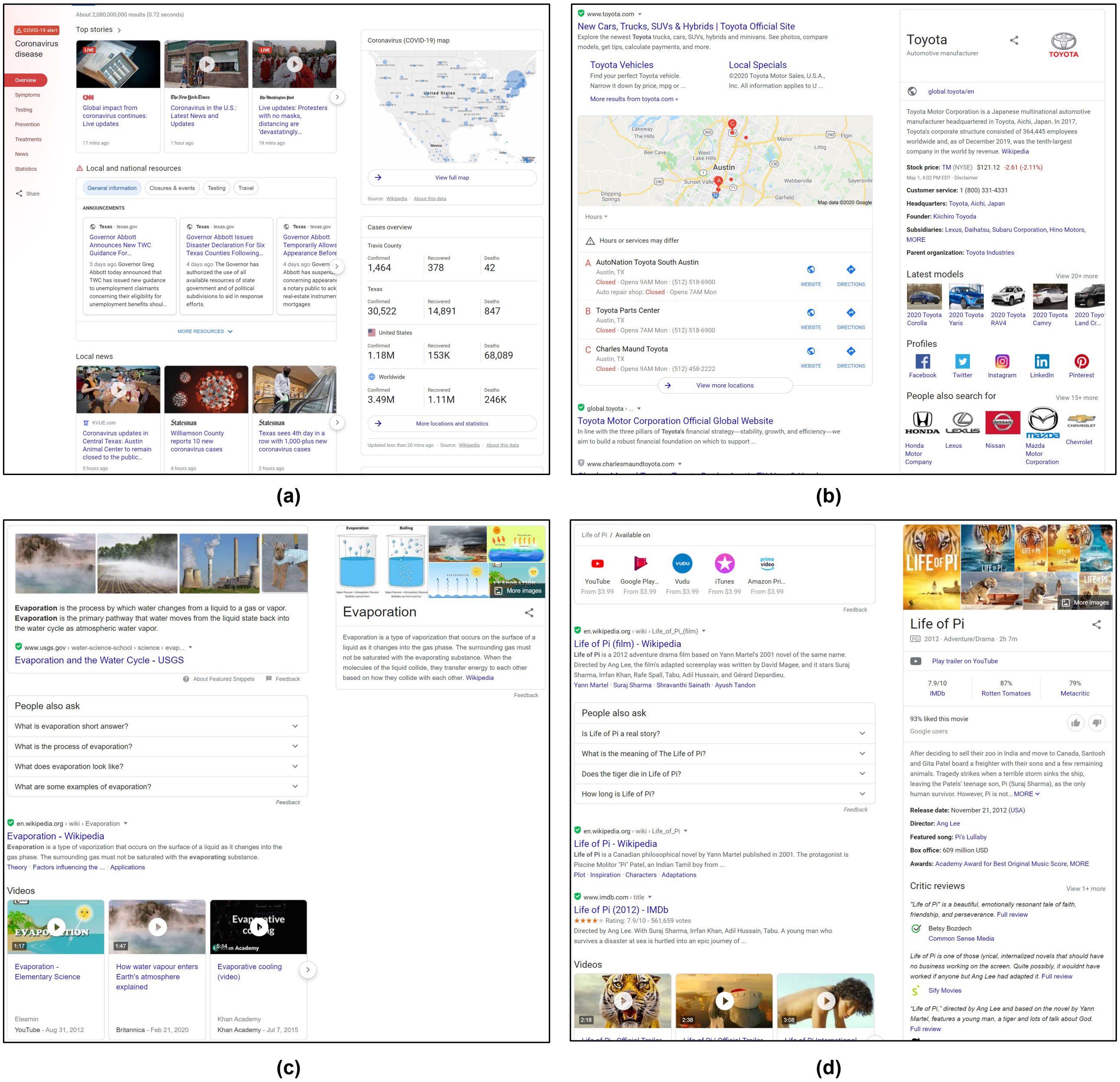


Figure 3.8: Google search engine result page (SERP) for the queries: **(a)** "coronavirus" **(b)** "toyota" **(c)** "evaporation", and **(d)** "life of pie". All screenshots are from 'above-the-fold', viewed on a 2560×1440 monitor. These examples highlight that modern SERPs have come a long way from a list of "ten blue links". SERPs are becoming consumable information-objects in their own right, and thus require different kinds of cognitive processing and interactions, than from the early days of the internet. Inspired and adapted from Wang et al. (2018). Accessed on May 5, 2020. This figure is best viewed in colour.

The final list contains 15 relevance criteria (e.g., *topicality*, *quality*, *recency*, *scope*, *availability*, etc.) and can be found in (Balatsoukas & Ruthven, 2012 Appendix B).

Search engines are increasingly adding different modalities of information on the SERP, besides the "ten blue links". These include images, videos, encyclopaedic information, and maps (Figure 3.8). Z. Liu et al. (2015) studied the influence of these different forms of SERP information – called 'verticals' – on searcher's relevance judgements. A general observation was that if verticals were present in a SERP, they created strong attraction biases. The attraction effect was influenced by the type of verticals, while the vertical quality (relevant or not) did not have a major impact. For instance, 'images' and 'software download' verticals had higher visual attention, while news verticals had equal attention as the "ten blue links" search results.

3.2.3 Stage 3: Content Page Evaluation / Item Examination

> *How do users behave when examining a single information-object (e.g., a a non-search-engine webpage, aka content page) obtained from an IR system?*

In online information searching, searchers repeatedly interact with individual webpages, a.k.a. 'content pages' in IR terminology. These webpages can be visited by following links from a search engine, following links between different webpages, or directly typing the URL in the browser.

The first group of papers we discuss investigated users' **visual attention** and **reading behaviour** on webpages. Pan et al. (2004) studied whether eye-tracking scanpaths on webpages varied based on task-type, webpage type (business, news, search, or shopping), viewing order of webpages, and gender of users. The found significant differences for all factors, except for task-type, which seemed to have no effect on scanpaths. They used weak task-types: remembering what was on a webpage vs. no specific task. In a later work on using informational vs. navigational search-tasks, they again saw limited effect of task-type on visual attention (Lorigo et al., 2006). Findings from Josephson & Holmes (2002)'s study suggested that users possibly follow habitually preferred scanpaths on a webpage, which can be influenced by factors like webpage characteristics and memory. However, they used only three webpages, making the findings difficult to generalize. Goldberg et al. (2002) studied eye movements on Web portals during search-tasks, and saw that header bars were typically not viewed before focusing the main part of the page. So they suggested placing navigation bars on the left side of a page. Beymer et al. (2007) focused on a very specific feature on webpages: images that are placed next to text content and how they influence eye movements during a reading task. They found significant influence on fixation location and duration. Those influences were dependent on how the image contents related to the text contents (i.e., whether they showed ads or text-related images). Buscher et al. (2009) presented findings from a large scale study where users performed information-foraging and page-recognition tasks. They observed that in the first few moments, users quickly scanned the top left of the page, presumably looking for clues about the content, provenance, type of information, etc. for that page. The elements that were normally displayed in the upper left third of webpages (e.g., logos, headlines, titles or perhaps an important picture related to the content) seemed to be important for recognizing and categorizing a page. After these initial moments, influence of task-type set in. For page-recognition tasks, the attention remained in the top-left corner of the webpage. However, for information-foraging tasks, fixations moved to the center-left region of the webpage, where the user was possibly trying to find task-specific information. The right third of webpages attracted almost no visual attention during the first one-second of each page view. Afterwards as well, most users seemed to entirely ignore this region, or only occasionally look at it. This suggested that users had low expectations of information-content or general relevance on the right side of most webpages. As many webpages display advertisements on the right side, this was a plausible observation, and are in line with the observed “F-shaped-patterns” [3] on webpages.

Buscher et al. (2009) also proposed an eye-tracking measure called *fixation impact*. This measure first appends a circular Gaussian distribution around each fixation on a webpage element, to create a fuzzy area of interest. This is called the *distance impact* value. If a webpage element completely covers the fixation circle (Gaussian distribution), it gets a *distance impact* value of 1. If the element partially covers the fixation circle, its *distance impact* value is smaller.

[3] https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content

Multiplying the *distance impact* value with the fixation duration gives the fixation impact for the given webpage element. Thus, an element that completely covers the fixation circle gets the full fixation duration as *fixation impact* value. Elements which are partially inside the circle get a value proportional to the Gaussian distribution. The authors posited that the rationale behind creating the fixation impact measure was motivated by observations from human vision research, which indicates that fixation duration correlates with the amount of visual information processed; the longer a fixation, the more information is processed around the fixation centre. Using the fixation impact measure, Buscher et al. (2009) proposed a model for predicting the amount visual attention that individual webpage elements may receive (i.e. visual salience).

Another group of studies investigated how users judged **relevance of webpages** w.r.t. an assigned search-task or information need. (Gwizdka, 2018; Gwizdka & Zhang, 2015a, 2015b) observed that when relevant pages were revisited, the webpages were read more carefully. Pupil dilations were significantly larger on visits and revisits to relevant pages, and just before relevance judgements were made. Certain conditions of visits and revisits also showed significant differences in EEG alpha frequency band power, and EEG-derived attention levels. Relevance of individual webpage elements were also assessed as *click-intention*: whether users would click on an element they were looking at. Slanzi et al. (2017) used pupillometry and EEG signals to predict whether a mouse click was present for each eye fixation. EEG features included simple statistical features of signals (mean, SD, power, etc.), as well as sophisticated mathematical features (Hjorth features, Fractal Dimensions, Entropy, etc.). A battery of classifier models were tested. However, the results were not promising. Logistic Regression had the highest accuracy (71%), but very low F1 score (0.33), while neural network based classifiers the had highest F1 score (0.4). The authors suspected that the low sampling rate of their instruments (30 Hz eye-tracker and 128 Hz 14-channel EEG) impacted their classifier performances. González-Ibáñez et al. (2019) compared relevance prediction performances in the presence and absence of eye-tracking data, and argued that when eye-tracking data collection is not feasible, mouse left-clicks can be used a good alternative indicator of relevance.

The '*Competition for Attention*' theory states that items in our visual field compete for our attention (Desimone & Duncan, 1995). Djamasbi et al. (2013) studied web search and browsing from the perspective of this theory. Theoretical models suggest that in goal-directed searches, information-salience and/or information-relevance drives search behaviour (i.e. competition for attention does not hold true), whereas exploratory search behaviour is influenced by competition among stimuli that attracts a user's attention (i.e. competition for attention holds true). However, in practice, information search behaviour often becomes a combination of both types of visual search activities (Groner et al., 1984). Djamasbi et al. (2013) found that, despite the goal directed nature of their search-task (finding the best snack place in Boston to take their friends) *competition for attention* had some effect at the content page level. Some of the users' attention was diverted to non-focal areas on content pages. However, there was little effect of *competition for attention* on how the results were viewed on SERPs. Users exhibited the familiar top-to-bottom pattern of viewing (Section 3.2.2), paying the most attention to the top two entries.

3.3 Effects of Expertise and Working Memory on Search Behaviour

Our focus of discussion in this dissertation is information searching and learning. As we saw in Chapter 2, learning and expertise are closely connected: expertise is an evolving characteristic of users that reflects learning over time, rather than being a static property (Rieh et al., 2016; Sawyer, 2005). (White, 2016a, Chapter 7) considers three types of expertise, that are relevant in information seeking settings: *(i)* domain or subject-matter expertise; *(ii)* search expertise; and *(iii)* task expertise. **Domain or subject-matter expertise** describes people's knowledge in a specialised subject area such as a domain of interest. **Search expertise** refers to people's skill level at performing information-seeking activities, both in a Web search setting and in other settings such as specialised domains. **Task expertise** describes people's expertise in performing particular search tasks, potentially independent of domain. Although considered distinctly, the boundaries between these expertise types are quite blurred, and therefore difficult to estimate at the time of search, and model it in a way that can be consumed by search systems.

Search Stage	Search behaviours indicative of learning, or increasing domain expertise
Query (re)formulation	- Increase in the *number* and *specificity* of query terms - Increase in number of synonyms - Decrease in number of reformulated queries
Search Engine Results Page (SERP) examination *(Source Selection)*	- Increased clarity in relevance criteria = increased ability to distinguish between relevant and non-relevant results - Decrease in the number of search results viewed (supported by Mao et al. (2018), contrasted by White et al. (2009)) - Decrease in the proportion of partially relevant results viewed, and increase in the number of relevant results viewed - Average time for assessing a search result decreases
Content Page examination *(Interaction with sources)*	- Increase in the amount of information-use from viewed content pages in the learning outcome artefact (summary, project report, exam answers, etc.) Knowledge Assimilation: addition of new information to existing knowledge structure - Focus on factual and specific information - Refining output with factual information - Revisiting content pages for information initially overlooked Knowledge Restructuring: large changes or replacement of concepts and their relations in knowledge structure - Focus on background and conceptual information; notes taken on themes and ideas - Ideas are related and combined for a focus, in the outcome Knowledge Tuning: small changes in scope and meaning of concepts and their relations in knowledge structure; no replacements - Focus on procedural and specific information - Identification of information to support and refine focus
Overall search session	- Decreased time per search session - Decrease in variability of search tactics - Increase in the diversity of websites visited within a subject area (increase in the average number of unique top-level websites on a SERP or across clicked documents) - Increase in focus of exploration (e.g., the degree to which a SERP is covered by a single topic) - Search path is more ‘branchy’ – returning to a previously visited point and then following a new unexplored direction) (White et al., 2009)

Figure 3.9: Literature reviews by Rieh et al. (2016) and Vakkari (2016) identified the following search behavioural traits as indicative of domain experts, or novices undergoing learning to become experts.

Previous work on domain knowledge and expertise have linked [4] domain expertise and search behaviour in terms of metrics, behavioural patterns, and criteria (M. J. Cole et al., 2013; Mao et al., 2018; O'Brien et al., 2020; White et al., 2009). A representative summary is presented in Figure 3.9, and is adapted from literature reviews by (Rieh et al., 2016) and (Vakkari, 2016). Briefly, (Wildemuth, 2004) showed that novices converge toward the same search patterns as experts, as they are exposed to a topic and learn more about it. (X. Zhang et al., 2011) found that features such as document retention, query length, and the average rank of results selected could be predictive of domain expertise. (M. J. Cole et al., 2013) showed that eye-gaze patterns could be used to predict an individual's level of domain expertise using estimates of cognitive effort associated with reading. (White et al., 2009) showed that measures such as diverse website visitation, more narrow topical focus, less diversity (or entropy), more 'branchiness' of search sessions, less dwell time, and higher query and session complexity are indicative of expert knoweldge and/or search behaviour.

As a stark contrast, (Zlatkin-Troitschanskaia et al., 2021) reviewed literature on higher education **students' information search behaviour**. Students can be considered as novices in all three respects: domain/subject-matter, search skills, and task. The authors report that across literature, higher education students' information search behaviour tends to follow some general general patterns: *(i) foraging:* no explicit (task-specific) research plan and little understanding of the differences (pros/cons) between various IR systems; *(ii) Google dependence:* no intention to use any search tool other than Google, causing students to struggle to understand library information structures and engage with scholarly literature effectively; *(iii) rudimentary search heuristic:* reliance on one and the same simple search strategy, regardless of search context; *(iv) habitual topic changing:* students change the search topic after rather superficial skimming, and before evaluating all search results; and *(v) overuse of natural language:* students type questions into the search box that are phrased as if posing them to a person. Highly ranked online sources accessed via a well-known search engine were perceived as trustworthy.

[4] and continue to link

Effects of memory span and working memory capacity have also been found to influence search effort and search behaviour (Arguello & Choi, 2019; Bhattacharya & Gwizdka, 2019a; L. Cole et al., 2020; Gwizdka, 2013, 2017). **Working memory** (WM) is considered a core executive function is defined as someone's ability to hold information in short-term memory when it is no longer perceptually present (Diamond, 2013; G. A. Miller, 1956). (Bailey & Kelly, 2011) showed that the amount of effort was a good indicator of user success on search tasks. (Smith & Kantor, 2008) studied searcher adaptation to poorly performing systems and found that searchers changed their search behaviors between difficult and easy topics in a way that could indicate that users are satisficing. Differences in search effort between different types systems (higher effort invested in searching library database vs. web) were found by (Rieh et al., 2012). A couple of studies showed that mental effort involved in judging document relevance is lower for irrelevant and higher for relevant documents (Gwizdka, 2014; Villa & Halvey, 2013). (Gwizdka, 2017) found that that higher WM searchers perform more actions and that most significant differences are in time spent on reading results pages. Behaviour of high and low WM searchers were also found to change differently in the course of a search task performance.

3.4 Assessing Learning during Search

In order for IR systems to foster user-learning at scale, while respecting individual differences of searchers, there is a need for measures to represent, assess, and evaluate the learning process, possibly in an automated fashion. Consequently, a variety of assessment tools have been used in prior studies. These include self reports, close ended factual questions (multiple choice), open ended questions (short answers, summaries, essays, free recall, sentence generation), and visual mapping techniques using concept maps or mind maps. Each approach has its own associated advantages and limitations. **Self-report** asks searchers to rate their self-perceived pre-search and post-search knowledge levels (Ghosh et al., 2018; O'Brien et al., 2020). This approach is the easiest to construct, and can be generalised over any search topic. However, self-perceptions may not objectively represent true learning. **Closed ended questions** test searchers' knowledge using factual multiple choice questions (MCQs). The answer options can be a mixture of fact-based responses (*TRUE*, *FALSE*, or *I DON'T KNOW*), (Gadiraju et al., 2018; Xu et al., 2020; Yu et al., 2018) or recall-based responses (*I remember / don't remember seeing this information*) (Kruikemeier et al., 2018; Roy et al., 2020). Constructing topic-dependant MCQs may take time and effort, since they are topic dependant. Recent work on automatic question generation may be leveraged to overcome this limitation (Syed et al., 2020). Evaluating close ended questions is the easiest, and generally automated in various online learning platforms. Multiple choice questions, however, suffer from a limitation: they allow respondents to answer correctly by guesswork. **Open ended questions** assess learning by letting searchers write natural language summaries or short answers (Bhattacharya & Gwizdka, 2018; O'Brien et al., 2020; Roy et al., 2021). Depending on experimental design, prompts for writing such responses can be generic (least effort) (Bhattacharya & Gwizdka, 2018, 2019b), or topic-specific (some effort) (Syed et al., 2020). While this approach can provide the richest information about the searcher's knowledge state, evaluating such responses is the most challenging, and requires extensive human intervention (Kanniainen et al., 2021; Leu et al., 2015; M. J. Wilson & Wilson, 2013) (as discussed in Section 2.4.2). **Visual mapping** techniques such as mind maps and concept maps have also been used to assess learning during search (Egusa et al., 2010, 2014a, 2014b, 2017; Halttunen & Jarvelin, 2005). Concept maps have been discussed at length in Section 2.3.1. Learning has also been measured in **other ways**, such as user's familiarity with concepts and relationships between concepts (Pirolli et al., 1996), gains in user's understanding of the topic structure, e.g., via conceptual changes described in pre-defined taxonomies (P. Zhang & Soergel, 2016), and user's ability to formulate more effective queries (Chen et al., 2020; Pirolli et al., 1996).

3.5 Limitations of Current Search Systems in Fostering Learning

3.5.1 Longitudinal studies

Learning is a longitudinal process, occurring gradually over time (Sections 2.3 and 2.2). Therefore, information researchers have studied participant's search behaviour in prior, **albeit few**, longitudinal studies. Examples include studies by (Kelly, 2006a, 2006b; Kuhlthau, 2004; Vakkari, 2001a; White et al., 2009; Wildemuth, 2004).

(Wildemuth, 2004) examined the search behaviour of medical students in microbiology. In this experiment, students were observed at three points of time (at the beginning of the course, at the end of the course, and six months after the course), under the assumption that domain expertise changes during a semester. Some search strategies, most notably the gradual narrowing of the results through iterative query modification, were the same throughout the observation period. Other strategies varied over time as individuals gained domains knowledge. Novices were less efficient in selecting concepts to include in search and less accurate in their tactics for modifying searches. (Pennanen & Vakkari, 2003; Vakkari, 2000, 2001a, 2001b) also examined students at multiple points in time, as they were developing their thesis proposal. One important change in behaviour was the use of a more varied and more specific vocabulary as students learned more about their research topic. (Weber et al., 2019) examined a large sample of German students from all academic fields in a two wave study and found that the more advanced they are in their studies, the more students show a more advanced search behaviour (e.g., using more English queries and accessing academic databases more frequently). **Advanced search behaviour predicted better university grades.** (Weber et al., 2018) also provide mixed evidence on the potential long-term effects of such interventions, as some of their participants reverted to their previous habits two weeks after the study and therefore exhibited only short-term changes in their information-seeking behaviour.

Overall, results regarding the promotion of user' search and evaluation skills are encouraging. But there is a clear need for more longitudinal studies. The general body of search-as-learning literature examines the learner in the short-term, typically over the course of a single lab session (Kelly et al., 2009; Zlatkin-Troitschanskaia et al., 2021). The trend is similar in other Human-Computer Interaction (HCI) research venues. A meta-analysis of 1014 user studies reported in the ACM CHI 2020 conference revealed that more than 85% of the studies observed participants for a day or less. To this day, "longitudinal studies are the exception rather than the norm" (Koeman, 2020). "An over-reliance on short studies risks inaccurate findings, potentially resulting in prematurely embracing or disregarding new concepts" (Koeman, 2020).

3.5.2 Supporting sensemaking and reflection

As we saw in Section 2.3, learning *is* sensemaking. Yet, modern search systems are still quite far from supporting sensemaking and learning, and rather, at best are good *locators* of information. (Rieh et al., 2016) says that modern search systems should support sensemaking by offering more interactive functions, such as tagging for annotation, or tracking individuals' search history, so that a learner could return to a particular learning point. In addition, a system could provide new features that allow learners to reflect upon their own learning process and search outcomes, thus facilitating the development of critical thinking skills.

> *It's easy to be impressed by the scientific and engineering feats that have produced web search engines. They are, unquestionably, one of the most impactful and disruptive information technologies of our time. However, it's critical to remember their many limitations: they do not help us* ***know what we want to know****; they do not help us* ***choose the right words*** *to find it; they do not help us know if what we've found is* ***relevant or true****; and they do not help us* ***make sense of it****. All they do is quickly retrieve what other people on the internet have shared. While this is a great feat, all of the content on the internet is far from everything we know, and quite often a poor substitute for expertise.*
>
> — Ko (2021) (emphasis our own)

3.6 Summary

In this second chapter of the background literature review, we discussed *(i)* how searchers interact with three stages / interfaces of modern information retrieval system: query formulation, search results evaluation, and content page evaluation; *(ii)* how expertise and working memory influence overall search behaviour; *(iii)* how learning or knowledge gain during search has been assessed in recent search as learning literature; and *(iv)* what are the limitations of current search systems to foster learning, including gaps in literature about long term search behaviour and learning outcomes, as well as lack of support for sensemaking.

We saw that while we have a plethora of studies investigating search behaviour searchers in the short term, we have merely a handful of studies observing the same participant for more than a day. To the best of the author's knowledge, most of these studies were conducted over a decade ago. Thus, while we have excellent knowledge of short term nature of influence of searching on learning, we do not know what are the longer term effects. Furthermore, we we have gaps in our knowledge of *(i)* how practices like articulation and externalization, and user attributes like metacognition, motivation, and self regulation moderate the searching as learning process; *(ii)* how these moderator variables change over time; and *(iii)* what these phenomena collectively entail for the design of future learning-centric IR systems. In the next chapter, we take these gaps in knowledge and use them to inform our research questions and hypotheses.

4

Research Questions and Hypotheses

4.1 Research Questions

Combining empirical findings and gaps in the literature from the disciplines of Education (Chapter 2) and Information (Chapter 3), we saw that:

- searching for information online is an integral part of new learning (Section 2.4.3)
- learning happens when students connect new pieces of information to their existing knowledge structures via assimilation, restructuring, or tuning (Section 2.3), and this process is influenced by the learner's individual traits (Section 2.5)
- modern knowledge-work requires less of long term memory, and more of creation of knowledge-artefacts, which should be treated as better assessors and outcomes of learning (Section 2.4.2)
- domain expertise and search behaviour are strongly linked (Section 3.3)
- learning is a process that takes place longitudinally over time (Sections 2.3 and 2.2), yet only a handful of studies (mostly over a decade ago) have investigated the intertwined process of searchers' learning and their information searching behaviour over time (Section 3.5.1)

- this creates acute gaps in our knowledge about long term information searching and learning behaviour, which is crucial for building learning-centric search systems of the future, which can support sensemaking and knowledge-gain

Guided by the above insights, we asked the following research questions in this dissertation, and aimed to answer them via a longitudinal study of students' information search behaviour and learning outcomes over the course of a university semester (Section 5.1). For the purposes of this dissertation, we considered learning as change in a student's knowledge about certain topics over the duration of a university semester.

The research questions are first stated in this section, to put them all together in one place for easy reference. Then the overarching hypotheses are discussed in Section 4.2.

RQ1: *What kind of longitudinal information search behaviours are correlated to the degree of change in students' knowledge levels and learning outcomes?*

RQ2: *What are the similarities and differences in information search behaviours for tasks where the learning goals are new (new search tasks), versus those where the learning goals are repeated (repeated search tasks)?*

RQ3: *How does externalisation and articulation affect students' learning outcomes and experiences during search?*

RQ4: *How do (changing) individual differences of students moderate their information search behaviours and learning outcomes?*

4.2 Overarching Hypotheses

In this Section, we discuss the research framework and hypotheses behind the research questions. The study is primarily planned to be exploratory, therefore the hypotheses are exploratory in nature as well.

4.2.1 Learning as Students' Transition from Novice to Expert (RQ1, RQ2)

Learning and expertise are closely connected: expertise is an evolving characteristic of learners that reflects learning over time, rather than being a static property (Rieh et al., 2016). Domain expertise and search behaviour has been studied, albeit mostly during single lab sessions, and sometimes longitudinally (Section 3.3). There is a clear gap in understanding how higher education students search for information in the long term, how their information use behaviour develop over time, and how it affects their learning (Zlatkin-Troitschanskaia et al., 2021). RQ1 and RQ2 aims to address some of these gaps.

Hypothesis for RQ1: Search behaviours described in Table 3.9 will occur both within individual search sessions, and across progressive search sessions recorded over a semester, as domain expertise of students increases (Eickhoff et al., 2014).

Hypotheses for RQ2: This research question stems from the idea of lifelong or continuous learning: how do search behaviours evolve over time when gaining knowledge about perpetual life skills (e.g., financial literacy). We hypothesize that

- relevance judgement of previously viewed information on this topic will change over time, as searcher gains more knowledge and expertise
- the decision or choice to put effort into searching again, or satisfice with previously found information, will have links to motivation and self-regulation

4.2.2 Promoting Better Learning (RQ3, RQ4)

Better learning takes place when students articulate and their unformed and still developing understanding, and continue to articulate it throughout the learning process (Section 2.5.1). Also, students' motivation, self-regulation and metacognition capabilities determine, direct, and sustain the approaches they take to learn (Section 2.5). Effective searching for learning is affected by students' search tactics and information evaluation capabilities (Section 2.4.3) as well as cognitive capabilities such as memory span (Section 3.3).

Hypothesis for RQ3: articulation during the search as learning process (via concurrent think aloud) will lead to better learning (and possibly better searching) outcomes, than working silently.

Hypotheses for RQ4: with respect to the individual differences and contexts in which students search to learn, we speculate the following hypotheses:

- students showing sustained or increasing metacognition, self-regulation, and motivation over the duration of the semester will put more "effort" into their searches, and demonstrate better learning and search outcomes

- students with higher memory span will demonstrate more 'branchiness' and parallel browsing in their search behaviour

- students with better information evaluation capabilities will demonstrate better learning and search outcomes

4.3 Anticipated Contributions

We anticipate by answering the proposed research hypotheses and question, the results can greatly contribute to the existing knowledge of Interactive Information Retrieval and Educational Sciences in general, and Search as Learning in particular. Referring back to some of the research agenda advocated by the multiple workshops and journal special issues on Search as Learning (Section 1.4), our research questions aim to investigate *(i)* the contexts in which students search to learn; *(ii)* the factors that influence their learning outcomes; and *(iii)* whether students are more critical consumers of information.

Many researchers have expressed their concern with the lack of longitudinal studies in IIR and related domains (Kelly et al., 2009; Koeman, 2020; Zlatkin-Troitschanskaia et al., 2021). If significant relationships were to be found between students' information search behaviours and learning outcomes, the results of this dissertation can provide great insights and contributions towards *(i)* understanding how search behaviours can predict learning outcomes; *(ii)* creating reliable measures, methods, and instruments for capturing changes in people's knowledge level, learning experiences, and learning outcomes (Rieh, 2020); and *(iii)* developing search systems that better support learning and sensemaking.

5

Methods: LongSAL - the Longitudinal Study

To investigate the research questions and hypotheses discussed in Section 4.1, we conducted the LongSAL study. The following sections discuss the study design, apparatus and procedure.

5.1 Study Design and Participants

LongSAL (Longitudinal Search as Learning study) is a remote, exploratory, longitudinal study that was conducted between January and June 2022 (Spring semester) at the School of Information, University of Texas at Austin (UT Austin). The study was approved by The University of Texas at Austin Institutional Review Board (Submission ID: `STUDY00002136`, Date Approved: December 8, 2021).

Participants were recruited from the student pool enrolled in the required undergraduate core-course: *Ethical Foundations for Informatics* (Fleischmann et al., 2022). 18 participants originally signed up for the study; 10 participants fully completed all the phases of the study, and the remaining 8 dropped off at different points during the semester. Students enrolled in the course had to submit a research paper of 2,000-2,500 words as the final project for the course. There were has four checkpoints spread across the semester to submit the drafts in progress: (i) paper proposal, (ii) outline, (iii) rough draft, and (iv) final paper. Writing the research paper required choosing an informatics ethical dilemma, and applying three ethical perspectives covered in the course to explore potential solutions to the selected dilemma. This involved searching and navigating information online, finding at least 20 relevant external sources, combining ideas, and weaving a narration around the information found in the selected sources.

The study design was informed by running a pilot study during Summer 2021 semester, in partnership with two courses at UT Austin School of Information: *Information in Cyberspace*, and *Academic Success in the Digital University*. More details of the pilot study are presented in Appendix **??**.

5.2 Apparatus

5.2.1 YASBIL Browsing Logger

The YASBIL browsing logger (Bhattacharya & Gwizdka, 2021) was utilised for this study. YASBIL (Yet Another Search Behaviour and Interaction Logger)[1] is a two-component logging solution for ethically recording a user's browsing activity for Interactive IR user studies. It was developed by the author in early Spring 2021, and was employed in the pilot study for data collection and testing. YASBIL comprises a Firefox browser extension and a WordPress plugin. The browser extension logs browsing activity in the participants' machines. The WordPress plugin collects the logged data into the researcher's data server. YASBIL captures participant's behavioural data, such as webpage visits, time spent on pages, identification of popular search engines and their SERPs, tracking mouse clicks and scrolls, and the order and sequences of these events. The logging works on any webpage, without the need to own or have knowledge about the HTML structure of the webpage. To protect the privacy of participants, the logger software can be switched on or off by the participant. Participants received regular reminders to turn YASBIL on only when they were searching for information related to the course.

YASBIL offers ethical data transparency and security for participants, by enabling them to view and obtain copies of the logged data, as well as securely upload the data to the researcher's

[1] https://github.com/LongSAL/yasbil

	ENTRY QSNAIRE [QSNR1]	INITIAL PHASE [PHASE1]	LONGITUDINAL TRACKING PHASE [PHASE2A, 2B, 2C, 2D]	MID-TERM QSNAIRE [QSNR2]	FINAL PHASE [PHASE3]	EXIT QSNAIRE [QSNR3]
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 QSNR1	Four different points over the semester	Semester mid-point	After last day of classes	After PHASE3
Where	Asynchronous	**Synchronous**: Remote	Async	Async	**Sync:** Remote	Async
What	Only in QSNR1: −Consent Form Repeated in QSNR2 and QSNR3: −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 **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 (PHASE2). − 2A: Proposal − 2B: Paper Outline − 2C: Rough Draft − 2D: Final Paper Participants also shared (anonymized) assignment submission	Similar to QSNR1, with repeated components	**Two search tasks:** one task-topic repeated from PHASE1, one new; same format as PHASE1 **Website comparison assessment** from Stanford History Education Group (SHEG).	Similar to QSNR2 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: Longitudinal study procedure.

server over an HTTPS connection. Although developed using the cross-browser WebExtension API [2], YASBIL currently works in the Firefox Web Browser. So participants were instructed to install Firefox and YASBIL on their machines when they volunteered to participate in the study.

5.3 Procedure

The longitudinal study consisted of six data collection components, as illustrated in Figure 5.1. They comprise three asynchronous **questionnaires** (`QSNR1`, `QSNR2`, `QSNR3`), two **remote synchronous study phases** over Zoom video conferencing software (`PHASE1`, `PHASE3`), and a set of four asynchronous **longitudinal tracking phases** (`PHASE2a`, `PHASE2b`, `PHASE2c`, `PHASE2d`). These phases are discussed in detail in the following sections.

[2] https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions/Build-a-cross-browser-extension

5.3.1 QSNR0: Recruitment Questionnaire (Appendix B.1)

Participants were recruited for the study via the recruitment questionnaire (QSNR0). The questionnaire contained questions about demographic information of the participant pool. The description of the study and the link to the questionnaire was posted in the Canvas Learning Management System used for the I303 course.

5.3.2 QSNR1: Entry Questionnaire

After recruitment, participants completed the entry questionnaire (QSNR1). The purpose of QSNR1 was to capture their individual-differences, or moderating variables, at the beginning of the semester. Details of the data captured in SUR1 are described below, with references to sections in the Appendix, where the full-text of the questionnaire can be found.

5.3.2.1 Consent Form (Appendix B.2.1)

The first page of QSNR1 was online consent form for participating in the study. Participants were able to proceed with the study once they provided informed consent.

5.3.2.2 Motivation (Appendix B.3)

Adapted from the *Intrinsic Motivation Inventory (IMI)* by (Ryan, 1982), which is a multi-dimensional measurement device intended to assess participants' subjective experience related to a target activity (the assignments for the course they are taking). The instrument assesses participants' interest/enjoyment, perceived competence, effort/importance, pressure/tension, perceived choice, and value/usefulness, while performing a given activity, thus yielding six sub-scale scores. The pressure/tension and the perceived choice components were not included in the entry questionnaire QSNR1, and were present in the mid-term (QSNR2) and exit (QSNR3) questionnaires.

5.3.2.3 Self-regulation (Appendix B.4)

Adapted from the *Self-Regulation Questionnaire (SRQ)* by (J. M. Brown et al., 1999), which assess seven self-regulatory processes through self-report: receiving relevant information, evaluating the information and comparing it to norms, triggering change, searching for options, formulating a plan, implementing the plan, and assessing the plan's effectiveness (Section 2.5.4).

5.3.2.4 Metacognition (Appendix B.5)

Adapted from the *Metacognivite Awareness Inventory (MAI)*, originally proposed by (Schraw & Dennison, 1994) as a 52-item true / false questionnaire, and later revised by (Terlecki & McMahon, 2018) to use five-point Likert scales. The instrument measures two components of cognition through self-report: knowledge about cognition, and regulation of cognition (Section 2.5.2).

After completing QSNR1 offline, participants were instructed to prepare for the initial synchronous phase, `PHASE1`, by installing Firefox web browser and the YASBIL extension on their machines. This was a one-time step. If a participant could not find the time for this step, they were informed that an extra 5-10 minutes would be taken in the beginning of `PHASE1` to complete this step.

The entry questionnaire and the software installation took about 10-15 minutes to complete. Participants were compensated with USD 5 for their time for completing this step. The questionnaire was published to the I-303 course students in the first week of the Spring 2022 semester.

5.3.3 `PHASE1`: Initial Phase

The `PHASE1` of the data collection took place in the beginning of the semester. The data-collection took place over a Zoom video call combined with YASBIL browsing logger installed in the participants' machines. Participants were asked to share their screen for the whole duration of the phase. Their screens and audio were recorded for the entire duration. They had the freedom to turn off their video. The total time for `PHASE1` was expected to not exceed

1.5 hours (90 minutes). Participants were compensated with USD 25 for this phase. The different components of `PHASE1` are described below.

5.3.3.1 Training Search Task

Participants performed a training search task to familiarize themselves with how to operate the YASBIL browser extension to log their browsing activity. The training task took around 2-5 minutes.

5.3.3.2 `PHASE1-FINANCE` and `PHASE1-UBUNTU`: Two Actual Search Tasks

Participants performed two search tasks: `PHASE1-FINANCE`, and `PHASE1-UBUNTU`. The `PHASE1-FINANCE` task was repeated at the end of the semester as `PHASE3-FINANCE` task. The `PHASE1-UBUNTU` task was not repeated, and instead the `PHASE3-BIAS` task tooK its place. This helps to answer the research question RQ2 (Section 4.1). The order of the two search tasks were randomized.

The repeated search task `FINANCE` was on the topic of financial literacy, a topic that we posit can be considered as universally important to college students, and part of lifelong learning. The prompts for the `PHASE1-FINANCE` and `PHASE3-FINANCE` tasks are presented in Figure 5.2. The non-repeated search tasks were on topics that were taught in the I303 course: Ubuntu ethics (for `PHASE1`) and Algorithmic Bias (for `PHASE3`). The prompts for these tasks are present in Figure 5.3.

To answer RQ3 (effect of externalization and articulation in learning), each participant performed one of the search tasks while thinking aloud (Concurrent-Think Aloud or *CTA condition*), and performed the other search task in silence (*silent condition*). The choice of the search task for each of the conditions was randomized and balanced.

Each search task began with a pre-task questionnaire (Appendix C.1), which asked participants to self-rate their pre-search knowledge-level and interest on the topic. Then participants turned on the YASBIL browsing logger and started searching. The deliverable for each search task was a written summary (artefact). After participants are satisfied with

Repeated Search Task: *Financial Literacy*	
Prompt for Initial Phase [PHASE1-FINANCE]	**Prompt for Final Phase [PHASE3-FINANCE]**
Money management and financial literacy are essential life skills, and what better time to learn about them than in college? Write a note to your future self, about essential money-related advice and skills that college students should know and practice. **What to do**: • Find **at least three (3)** unique, good quality online resources that are relevant to this topic • Look for resources that help establish connections and develop a narrative **What to deliver**: • Write a summary of the lessons, advice, and/or tips you found across the different resources. This is a note to your future self, so the narrative can be in a format that is most useful and interesting to YOU :) • Paste the links of ALL the resources that you finally selected to develop your narrative, in the second text box, one link per line Please **turn YASBIL ON** before starting to search. Write the summary (note to your future self) here: _____ Paste the links of the resources that you finally selected (at least 3), one per line: _____	At the start of the semester, you wrote a note to your future self, about essential money-related advice and skills that college students should know and practice. Here is what you wrote: [… *dynamic content showing participant's* PHASE1 *response* …] Here are the resources you took help from: [… *dynamic content* …] Now you have a chance to **update or revise the note with more information**. You can either choose to write afresh, or copy-paste the note from above into the first textbox below and add to it /edit it. Feel free to search the web if you need to, after turning YASBIL on . You can choose **NOT to search**, as well. If you do choose to search, please paste the links of ALL the resources that you finally selected for updating your note, one link per line, in the second textbox. The links can be the same ones you visited earlier, or different. Please **turn YASBIL ON** before starting to search (if you choose to search). Write the updated summary (note to your future self) here: _____ Paste the links of the resources that you used to update your note here, one per line. If you did not search this time, type "N/A": _____ Did you need to search the web for updating the note? Why? Yes: _____ No: _____ Other: _____

Figure 5.2: Prompts for the search task that was repeated in the final phase, on the topic of financial literacy.

the quality of the deliverable, they turned off YASBIL browsing logger, and proceeded to the post-task questionnaire.

The post-task questionnaire (Appendix C.2) asked participants to self-rate their post-search topic knowledge, search experience, interest and motivation, and overall perceptions. The pre-task and post-task questionnaires are adapted from (Collins-Thompson et al., 2016; Crescenzi, 2020).

After the two search tasks were completed, participants answered questions about whether

Non-Repeated Search Tasks	
Prompt for Initial Phase: Ubuntu Ethics [PHASE1-UBUNTU]	**Prompt for Final Phase: Algorithmic Bias [PHASE3-BIAS]**
The term Ubuntu originates in Africa, and refers to the idea that our individual lives are intricately tied to the lives of others, and we all need to cherish these interconnections. Such interconnections should be guided by kindness, openness, accommodation, and willingness to work for others' interest. Do you think the Ubuntu philosophy can be used to mitigate the social, physical, and mental isolation that people may be facing during the COVID-19 pandemic? **<u>What to do</u>:** • Find **<u>at least three (3)</u>** unique, good quality online resources that are relevant to this topic • Look for resources that help establish connections and develop a narrative **<u>What to deliver</u>:** • Write a short summary of the content that you found across the different resources. • In the summary, briefly mention your thoughts about each resource - do you agree or disagree with the content in the resource? Anything else? • Paste the links of ALL the resources that you finally selected to develop your narrative, in the second text box, one link per line Please **turn YASBIL ON** before starting to search. Write the summary here: _____ Paste the links of the resources that you finally selected (at least 3), one per line: _____	In the I303 course, you studied about **Algorithmic Bias and its implications**. Therefore, for answering the questions below, **<u>you may choose NOT to search the web</u>**, if you feel you can answer the questions reasonably well. If you need to search the web, feel free to do so, **after turning YASBIL ON** . Paste the links of ALL the resources that you finally selected to develop your narrative, in the second text box, one link per line Please **turn YASBIL ON** before starting to search (if you choose to search). As you understand these concepts, briefly explain (with examples if necessary) 1. What is algorithmic bias? 2. How does algorithmic bias originate? 3. What are some implications of algorithmic bias? Response: _____ Paste the links of the resources that you used to write your answers, one per line. If you did not need to search, type "N/A": _____ Did you need to search the web for this task? Why? Yes: _____ No: _____ Other: _____

Figure 5.3: Prompts for the non-repeating search tasks. Topics were selected from the I-303 course content.

they preferred the think-aloud condition or the silent condition, and why (Appendix C.3).

5.3.3.3 `PHASE1-SHEG`: Website Reliability Assessment

To assess participants' (mis)information evaluation capabilities, participants performed a website reliability assessment created by the Stanford History Education Group (SHEG) (2021b).

> *This task presents students with the website of the American College of Pediatricians (ACPeds.org) and asks them whether it is a trustworthy source to learn about children's health. Despite the site's professional title and appearance, the American*

> *College of Pediatricians is not the nation's major professional organization of pediatricians. That designation belongs to the similarly named American Academy of Pediatrics.*
>
> *This exercise is an open web search in which students are free to stay on the American College of Pediatricians site or leave it to search for information about the group. Successful students will look beyond the surface features of the site and detect its agenda from its new releases or other focus issues. A faster route, however, is to leave the site almost immediately to search for reliable information about the true agenda of this organization.*
>
> — Stanford History Education Group (SHEG) (2021b)

The prompt for the `PHASE1-SHEG` task is present in Appendix C.4).

5.3.3.4 Memory Span Test

`PHASE1` concluded with the assessment of the participant's working memory capacity (WMC) using a memory span task (Francis et al., 2004). The task has 25 trials. On each trial participants saw a list of items presented one at a time in random order and were asked to recall the items in the same order in which they were presented. If they got a list correct, the list length increased by 1 for that type of material. If they got a list incorrect, the list length decreased by 1.

The type of material participants were asked to recall were: digits, letters that sound dissimilar, letters that sound similar, short words, and long words. The outcome score was the list length of the last list that participants could correctly recall.

5.3.4 `PHASE2A` - `PHASE2D`: Longitudinal Tracking Phase

The four-part longitudinal tracking phases `PHASE2A` - `PHASE2D` were conducted asynchronously over the duration of the semester, to understand the change (or lack thereof) of participants' search behaviour and knowledge gain over time. Whenever participants worked on different parts of their final project (Ethical dilemma research paper for the I-303 course), as described in Figure 5.4, they used Firefox web browser, and logged their browsing activity using YASBIL browsing logger. To protect their privacy, participants were regularly instructed to turn YASBIL on only when they were searching for information related to coursework. After each checkpoint

Final Project Description: Ethical Dilemma Research Paper

Throughout the semester, you will choose an ***informatics ethical dilemma*** and apply ***three ethical perspectives*** covered in the class to explore potential solutions to your selected dilemma. You are required to apply readings from the course as well as readings from outside of the course, including incorporating the three ethical theories as part of your analysis. This final project is broken down into several components that you will complete throughout the semester.

Proposal (week 3): must include the proposed ***title*** of your research paper as well as a one-page description of the ***informatics ethical dilemma***. At this stage, you are not required to have settled on your ethical theories yet. Please make sure to include your ***strategy*** for finding appropriate outside readings. **[PHASE2A]**

Outline (week 6): will typically include bulleted lists, filling in as much detail as you have ready at this point; must include the three planned ethical theories; must **list** of at least **10** potentially relevant references / scholarly readings, with **5** of them coming from beyond the course syllabus (external sources). **[PHASE2B]**

Rough Draft (week 10): must be at least a half-complete version of your final paper; should be 1,000-2,500 words; must **cite** at least **10** sources, including **5** from external sources. **[PHASE2C]**

Final Paper (week 15): must be complete, coherent, and easy to read; must incorporate feedback from all previous stages; should be 2000-2500 words; must cite at least **20** sources, including **10** external sources. **[PHASE2D]**

Peer Reviews of Rough Draft (week 10-11): peer-review the rough draft of two of your peers; prepare a one page peer review for each of them; must include: a brief summary of the purpose and content of the paper as you understood it, the strengths of the paper, and constructive feedback on how to improve it

Final Presentation (week 15): record a 5-min video presentation of your paper.

Figure 5.4: Final project description, setting up the longitudinal tracking phase of the study throughout the duration of the Spring 2022 semester. Text taken from I-303 course syllabus (Fleischmann et al., 2022); emphasis and annotations our own.

assignment, participants self-uploaded an anonymized version of the working-draft of their research paper, and answered a post-task questionnaire. The post-task questionnaire were similar to those used in the `PHASE1` and `PHASE3` search tasks, where participants self-reported, among other things, their perceived learning outcome and perceived search outcome (Collins-Thompson et al., 2016). Participants received reminder emails before the deadline of each assignment, to remind them to use Firefox, turn YASBIL on, and upload the anonymized working-draft. To prevent participant drop-off, a staggered payment model was adopted during PHASE2. Participants received USD 5 each when they completed `PHASE2A` and `PHASE2B`, USD 10 for `PHASE2C`, and USD 15 for `PHASE2D`, for a total of USD 35 for entire `PHASE2`.

5.3.5 `QSNR2`: Mid-Term Questionnaire

The mid-term questionnaire `QSNR2` took place around the mid-point of the semester (Week 8-9). The purpose was to track whether any of the participants' individual difference measures (motivation, metacognition, and self-regulation) changed during the first half of the semester. This questionnaire was essentially a replica of the Entry Questionnaire `QSNR1`, with two modifications. First, the consent form and the demographics sections were absent. Second, the Intrinsic Motivation Inventory (IMI) included the 'pressure/tension' and the 'perceived choice' subscales, as these scales are more meaningful after an activity has taken place (Ryan, 1982). The IMI was also be reworded to reflect the mid-point of the semester. Participants were compensated with USD 10 for completing this step.

5.3.6 `PHASE3`: Final Phase

The Final Phase `PHASE3` was similar in structure to the Initial Phase (`PHASE1`), and took place at the end of the semester, after all the course related tasks were completed by the participant. The purpose of the session is to record the 'evolved' search behaviour, and final knowledge state. Participants performed two search tasks: `PHASE3-FINANCE` and `PHASE3-BIAS`, and one website comparison task: `PHASE3-SHEG`.

5.3.6.1 `PHASE3-FINANCE` and `PHASE3-BIAS`: Two Actual Search Tasks

Of the two search tasks, the topic of one was repeated from `PHASE1` (financial literacy, Figure 5.2), while the topic of the other came from the course material: algorithmic bias (Figure 5.3). In both search tasks, participants were given the option of ***not searching*** if they felt confident enough to answer the search task questions from their prior knowledge (Crescenzi, 2020). The deliverables for each search-task, as before, was a written summary (artefact).

5.3.6.2 `PHASE3-SHEG`: Website Comparison Assessment

Following the two search tasks, participants performed another website comparison assessment from the Stanford History Education Group (SHEG) (2021a). This task assessed their evolved information evaluation skills.

> *This task instructed participants to compare two websites and select the one that they would use to begin research on a topic. One of the pages is a Wikipedia article. The other has ".edu" in its URL, but the page reveals that the content is a student-written blog post created as part of a university course. Many students have been taught that Wikipedia pages are unreliable and should be avoided. Many have also been taught that sites with a .edu domain are trustworthy. This assessment gauges their ability to think in more nuanced ways about these kinds of websites.*
>
> — Stanford History Education Group (SHEG) (2021a)

The prompt for the `PHASE3-SHEG` task is present in Appendix C.5).

Similar to `PHASE1`, participants were asked to share their screen for the whole duration of the phase. Their screen and audio was recorded for the same. They had the freedom to turn off their video. The total time for `PHASE3` was expected to not exceed 1.5 hours (90 minutes). Participants were compensated with USD 30 for `PHASE3`. At the end of `PHASE3`, participants were instructed to complete the Exit Questionnaire `QSNR3` as soon as convenient.

5.3.7 `QSNR3`: Exit Questionnaire

The exit questionnaire `QSNR3` took place after the Final Phase `PHASE3`. The purpose was to record the final state of the participants' individual difference measures (motivation, metacognition, self-regulation), and whether these characteristics changed during the second half of the semester. As before, `QSNR3` questionnaire was essentially be a replica of `QSNR2`, with the Intrinsic Motivation Inventory (IMI) reworded to reflect the end-point of the semester Participants were be compensated with USD 15 for their time for completing this step.

After `QSNR3` was complete, participants received a bonus compensation of USD 30, if they completed all the phases of the LongSAL study without missing anything.

5.4 Measures to Address Ethical Concerns

- Participation in the study (which was voluntary and compensated separately) and participation in the I303 course (which was required for graduation from the Informatics major) were sufficiently disentangled. The course instructors were never aware of which students participate in the course, and did not share any student data with the researchers. This avoided any undue pressure or expectation on the students.

- Participants logged their browsing activity using a Firefox browser extension YASBIL, which was been developed by the authors. The extension has an ON-OFF button, which put the participants in full control of when they wished to start and stop the logging. Participants had been sufficiently trained to use the browser extension, and were repeatedly reminded to log data only when they were working on the research paper assignments for the course, and not at other times.
- This study has been approved by The University of Texas at Austin Institutional Review Board (Submission ID: STUDY00002136, Date Approved: December 8, 2021).

After data collection for all the phases was complete, data analysis was performed on the collected data, which is discussed in the next chapter.

6 Data Analysis Framework

This chapter describes the general framework for analysing the data collected in the *LongSAL* study. It is described in Figure 6.1.

Here is the overall data analysis framework of my dissertation: Two categories of data were collected from participants. 1. Questionnaire data for individual differences: motivation, metacognition, self-regulation. 2. Search Log data containing URL visits, mouse events, webnavigation events, and the order and sequence of those events.

Questionnaire Data

1. Cleaning and Preprocessing
2. identifying user groups - latent profile analysis

Log Data

1. Log Data cleaning and pre-processing
2. URL categorization
3. Active Tab Identification
4. Higher levelSearch Behaviour Data Analysis framework - moves, tactics, strategies
5. Combining Latent Profiles with Search Behaviour Data

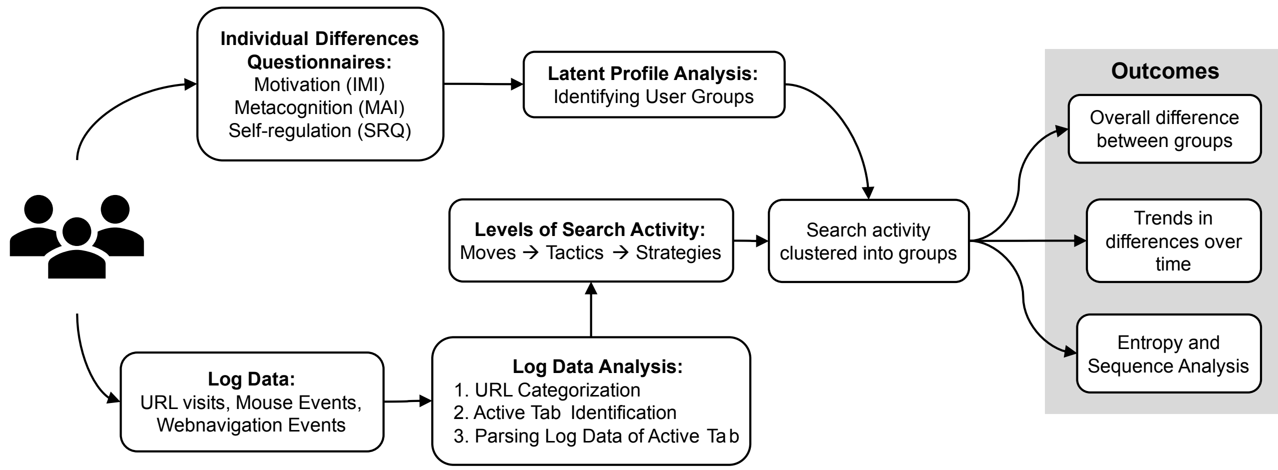


Figure 6.1: Data analysis framework followed in this dissertation.

Trends Over Time Difference Between Groups / Profiles Entropy analysis of sequences

6.1 Individual Differences Questionnaires Data Analysis Steps

6.1.1 Motivation, Metacognition, and Self Regulation Data

6.1.2 Latent Profile Analysis (LPA)

According to Ambrose et al. (2010), students' motivation, metacognition, and self-regulation are critical factors that determine, direct, and sustain what they do to learn. Given our interest in understanding how these traits impact students' searching as learning behaviour, we collected self-perceived reports of all three constructs, via the IMI, MAI, and SRQ questionnaires (Section 5.3.2). However, it is important to note that these constructs are not single binary variables that can be used to easily group individuals. Rather, they are complex and multidimensional data that serve as observable indicators of a person's underlying latent characteristics.

To cluster participants into meaningful groups based on these multiple constructs, we turned to the educational psychology literature. **Latent Profile Analysis (LPA)** is an increasingly popular statistical approach falling under the umbrella of person-centred techniques used in organizational psychology and child development research. It provides a framework for characterizing population heterogeneity in terms of differences across individuals on a set of behaviours or characteristics, as opposed to describing the variability of a single variable. By identifying latent subgroups within a population, LPA enables researchers to gain a more nuanced understanding of the complexity of human behaviour.

The person-centred approach underlying LPA is a departure from traditional variable-centred approaches such as multiple regression analysis. Instead of quantifying the role of particular variables in a study, LPA organizes a population into a finite number of mutually exclusive and exhaustive profiles, each comprising individuals who are similar to one another. In this way, LPA identifies distinct profiles of individuals who exhibit similar patterns of behaviour across multiple variables.

The identification and description of these latent profiles is a crucial step in LPA. Each profile represents a subgroup of individuals who share similar patterns of responses on a set of variables, which can provide insights into the underlying mechanisms driving their behaviour. Furthermore, the identification of the optimal number of profiles to represent a population is a critical issue in LPA. This involves balancing the complexity of the model with its ability to capture meaningful variability in the data, and requires careful consideration of both statistical and substantive criteria.

LPA has several advantages over traditional variable-centred approaches. It allows for a more nuanced understanding of the complexity of human behaviour, particularly in cases where individuals exhibit multiple and diverse patterns of behaviour across different sets of variables.

In the context of information search behaviour, LPA can help to identify distinct groups of individuals who engage in different search strategies or have different search motivations. This can be useful for understanding how people search for information online, what factors influence their search behaviour, and how search behaviour relates to other variables such as task performance, satisfaction, and learning outcomes.

The purpose of this study was to investigate the relationship between individual differences in motivation, metacognition, and self-regulation and search behaviour. To classify participants into high and low groups based on their scores on these questionnaires, we employed LPA.

LPA is particularly useful when the relationship between variables is not well understood or when it is difficult to determine which variables should be used to classify individuals into groups. We employed LPA to identify latent profiles of participants based on their scores on the IMI, MAI, and SRQ questionnaires.

6.2 YASBIL Search Log Data Analysis Steps

6.2.1 URL Categorization

- peer-reviewed publications are `PUB`s
- others are `ARTICLE`s (e.g. Wikipedia)
- if no other info, then `WEB`
- fuzzy between `WEB` and `ARTICLE` (when classified manually)
- `ARTICLE` if there is a clear author
  - except `WIKIPEDIA`, due to common parlance
  - encyclopedias
- journal homepages are `WEB`
- list of chapters in a book are `L.PUB`.
  - e.g. in detail view of
- book chapter is `PUB`

In order to understand the relationship between users' information search behavior and the type of webpages they visit, we needed to categorize webpages into different types. To accomplish this, we developed a classification system based on URL patterns.

URL patterns were first extracted from the web browsing data collected in our study. These URL patterns contain information about the structure and content of each webpage visited by the users. Based on this information, we were able to classify each URL present in the log data into the following hierarchical taxonomy:

- `L`: Search Result Pages, i.e., a **L**ist of Information Objects
  - `L.PUB`: Publication Search Results, e.g., on university library websites, digital libraries, Google Scholar, etc.
  - `L.WEB`: Web Search Engine Result Pages (SERPs)
- `I`: Content pages, i.e., Individual **I**nformation Objects
  - `I.PUB`: Academic Publications
  - `I.WEB`: Webpages that have the potential to provide relevant (academic) information for the search task, but are not publication. E.g., Wikipedia articles, relevant blog posts, government and non-profit websites, etc. Some of them were classified automatically (e.g., Wikipedia), while others were classified after manual inspection.
- `MISC`: URLs for webpages that did not fit in any of the above category

To identify search engine result pages, we looked for URLs that contained URL query parameters such as `q` (Google, Bing), `search`, `query`, or `k` (Yahoo) along with specific strings associated with popular search engines such as Google or Bing. We also identified content pages by looking for URLs that contained strings such as "`article`", "`blog`", or "`news`". Scientific peer-reviewed publications were identified based on URLs that contained specific strings associated with academic publishers or databases (ACM DL, Elsevier, Scopus, Springer etc.), while Wikipedia articles were identified by their URLs containing the string "wikipedia" in the hostname.

Library websites were identified by URLs that contained terms such as "library", "catalogue", or "database" as well as specific strings associated with major library systems (e.g., UT Austin uses Primo VE system from Ex Libris). Finally, we used the catch-all category `MISC` to identify other types of webpages that did not fit into any of the other categories.

Overall, our URL-based classification system provided a useful way to categorize webpages based on their type, allowing us to gain insights into how users' search behaviour varies across different types of webpages. By analysing the patterns of webpage types visited by users during their information search process, we were able to identify which types of webpages were most commonly visited and how they related to users' search behaviour. This information can be used to improve the design of information systems and search engines, as well as to inform the development of tailored interventions that support users' information search needs.

6.2.2 Active Tab Identification

6.2.3 Identifying Levels of Search Activity

After log data was cleaned From White (2016a), Table 2.1 (adapted from Bates, 1989):

- **Level 1: Move**
  - Atomic search event – for example, a query or click (*An identifiable thought or action that is part of information searching.*)
- **Level 2: Tactic**
  - Goal or task, including query or click chain (*One or several moves made to further a search*)
- **Level 3: Statagem**
  - Mission or session (*A larger, more complex set of thoughts and/or actions than the tactic; a stratagem consists of multiple tactics and/or moves, all of which are designed to exploit a particular search domain that is thought to contain the desired information*)
- **Level 4: Strategy**
  - Session or cross-session search task (*A plan, which may contain moves, tactics, and/or stratagems, for an entire information search.*)

Session: `task_id`, 30 minutes of inactivity (Google Analytics: a session lasts until there's 30 minutes of inactivity)

Session duration is considered 30 minutes (!) (or maybe 1 hour?) as per Google Analytics[1] and (TODO: find reference)

6.3 Combining Individual Differences with Search Logs

6.4 Entropy Analyses of Search Behaviour Sequences

Entropy is a measure of the diversity or unpredictability of a sequence of events. In the context of search queries, entropy can be used to quantify the variability or randomness of the query reformulations issued by participants. Transition analysis and entropy helps to cover differences in disparate tasks and activities. Inspired by previous works in analysing eye-movement sequences (Krejtz et al., 2014, 2015) and search tactics sequences (He et al., 2016), we employed a similar entropy analysis of query reformulation sequences, and search tactic patterns of the participants.

For query reformulations, the possible set of states were the five query reformulation types: Generalization, Specialization, Word Substitution, Repeat, and New.

For search tactic sequences, we used the following set of states:

1. `QUERY`: issuing a search query
2. `CLICK`: mouse click
3. `IDLE`: participant stays idle for more than one minute (TODO CITATION)
4. `SESSION_BREAK`: participant becomes idle for more than 30 minutes (Google Analytics defines a session break as 30 minutes of inactivity [2])
5. `L.PUB`: visiting a publication search result page
6. `L.WEB`: visiting a web SERP
7. `I.PUB`: visiting a scholarly publication
8. `I.WEB`: visiting a non-scholarly content page
9. `TASKPAGE`: visiting webpages related to the study, i.e. the Qualtrics questionnaires

[1] https://www.hotjar.com/google-analytics/glossary/session-duration
[2] https://support.google.com/analytics/answer/2731565#zippy=%2Cin-this-article

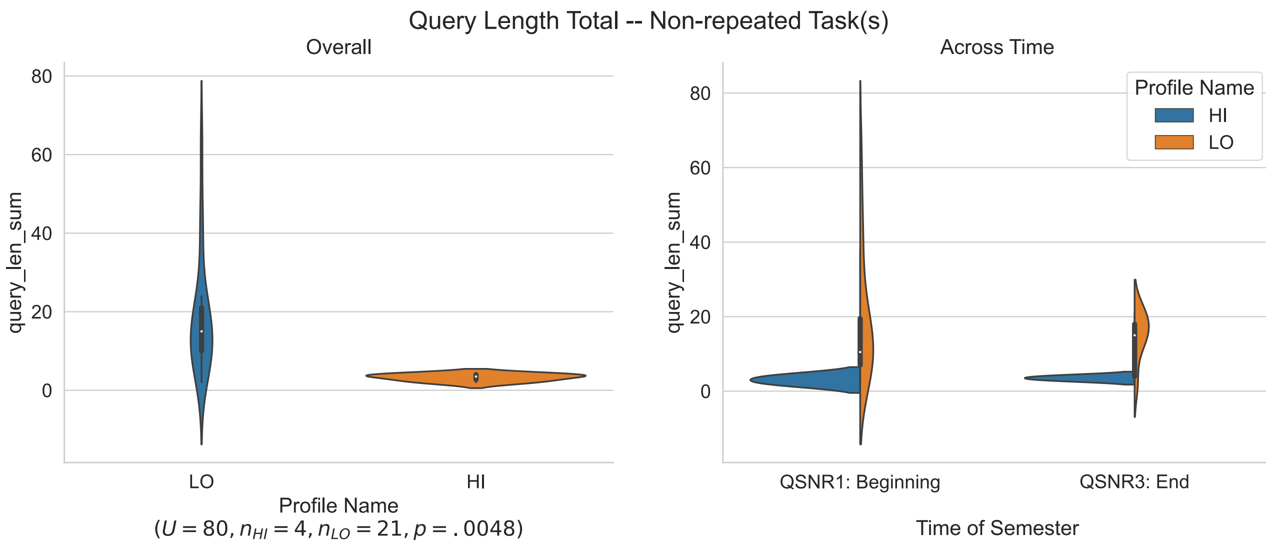


Figure 6.2: Longcaption.

If we consider a sequence of query reformulations issued by a participant (e.g., *New* -> *New* -> *Specialization* -> *Specialization* -> *Word Substitution* -> *Generalization*) then this sequence can be considered as a first order Markov chain, wherein, the next step in the chain depends only on the current state. Entropy analysis on these Markov chains quantifies how predictable the states are, and yields two categories of uncertainty measures: transition entropy, H_t, and stationary entropy H_s. Similar stationary and transition entropy measures can be obtained for sequences of search tactics.

7

Results: Longitudinal Tracking Phase

longitudinal study (cohort study)

Most of the results did not have statistical significance - but that's okay because this is exploratory research, which is inductive, rather than deductive. https://youtu.be/URz4CAaJqnE

We were more interested in discovering interesting patterns. We believe the findings from this study will lead to more confirmatory studies.

https://www.youtube.com/playlist?list=PLZp7Vke_WTVr8E4dM1fDeMJr2B7My9Gan

All statistical tests are two-tailed Mann-Whitney U test.

As the sample sizes were often small, the groups were imbalanced, and / or the data did not match the normality assumptions of parametric tests, we employed the non-parametric Mann-Whitney U test for all the null-hypothesis significance testing reported in this and the next chapter. This allows easy comparison between different categories of results.

Additionally, we also report Common Language Effect Size (CLES) for each statistical test result. The common language effect size is the proportion of pairs where x is higher than y. In other words, it is the probability that a score sampled at random from the first distribution will be greater than a score sampled from the second distribution. CLES was first introduced by McGraw and Wong (1992) [TODO]. The Python statistical library employed in data analysis –

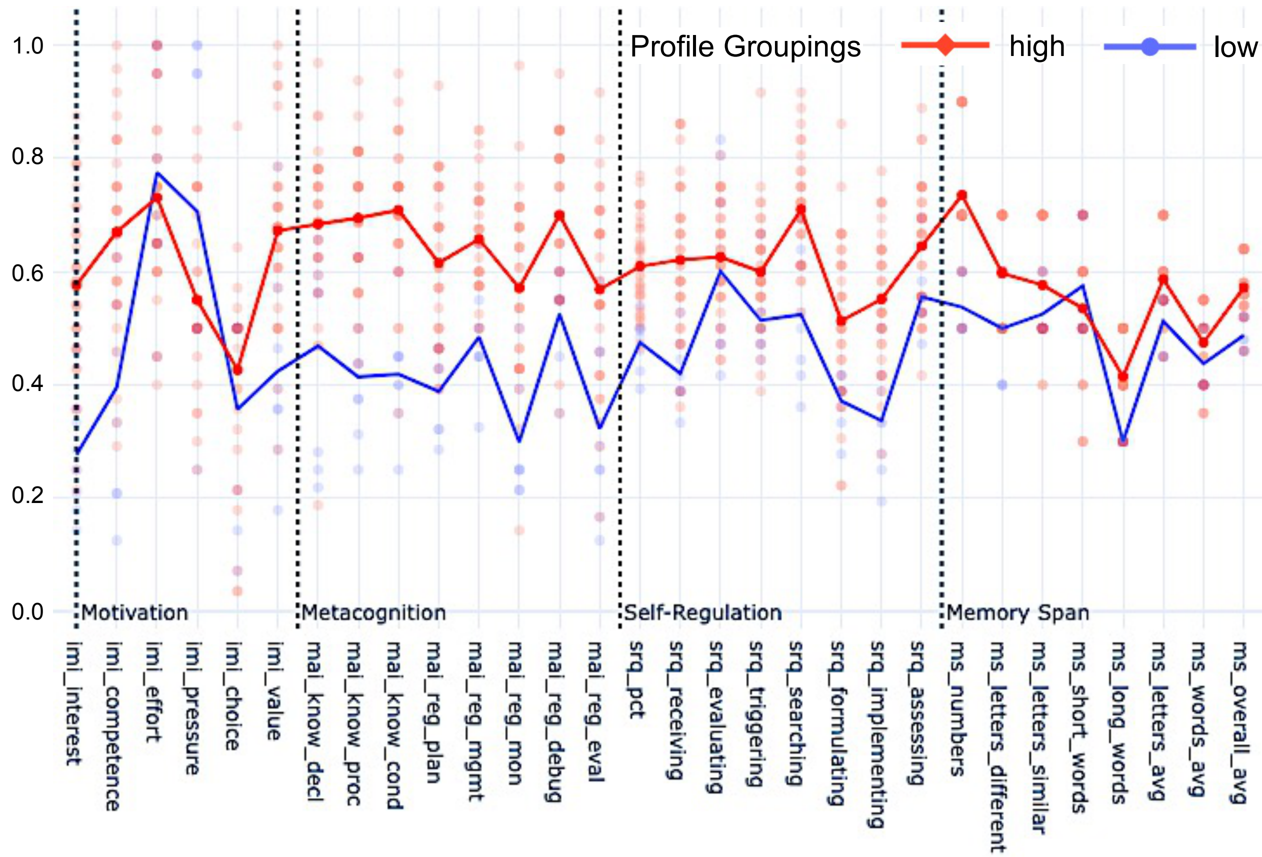


Figure 7.1: Mean Values of Indicator Variables for the two identified groups via Latent Profile Analysis (LPA). The grouping was based on the self-reported values of motivation (IMI), metacognition (MAI), self-regulation (SRQ), and a Memory Span task.

Pingouin (Vallat, 2018) – uses a brute-force version of the formula given by Vargha and Delaney 2000 [TODO]: The advantage is of this method are twofold. First, the brute-force approach pairs each observation of x to its y counterpart, and therefore does not require normally distributed data. Second, the formula takes ties into account and therefore works with ordinal data.

7.1 Latent Profile Analysis

We employed Latent Profile Analysis (LPA) to identify groups (latent profiles) of participants based on their scores on the Intrinsic Motivation Inventory (IMI), Metacognitive Awareness Inventory (MAI), Self-regulation questionnaire (SRQ), and Memory Span (MS) task. LPA is particularly useful when the relationship between variables is not well understood, or when it

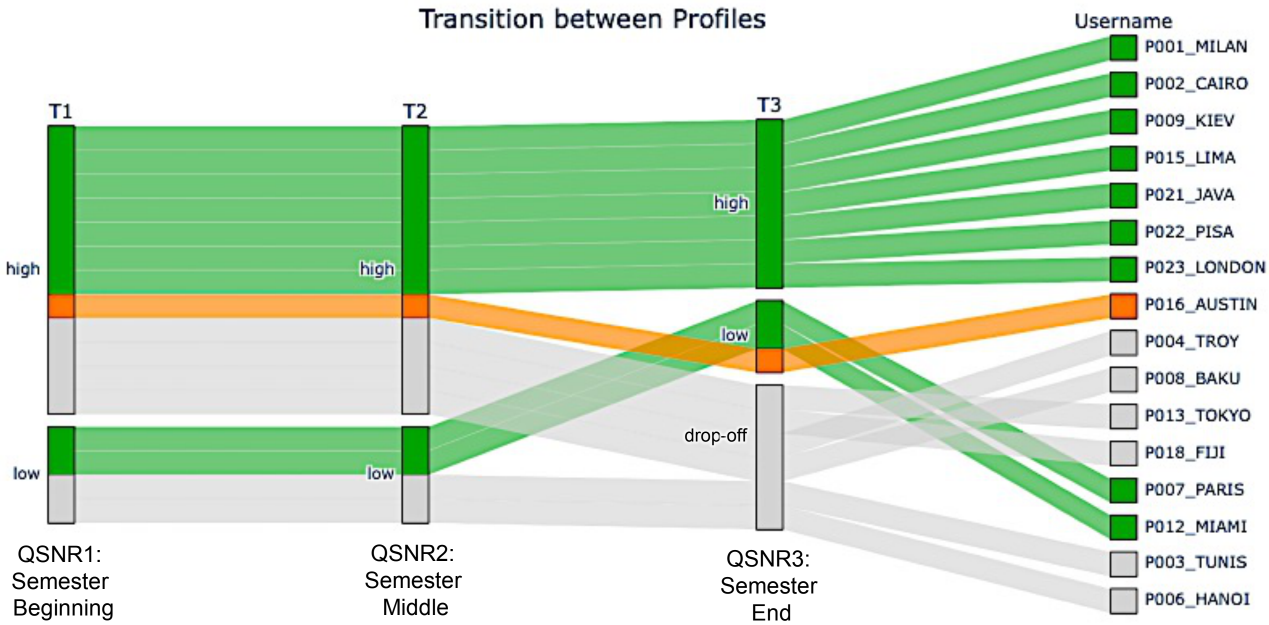


Figure 7.2: Diagram illustrating, at different timepoints, how participants stayed within their same high/low LPA profiles (green) or changed profiles (orange). Grey trajectories indicate participants who dropped off and did not complete the study.

is difficult to determine which variables should be used to classify individuals into groups.

The results of the LPA showed that there were two distinct groups (latent profiles) of participants based on their scores on the IMI, MAI, SRQ, and MS: a high group and a low group (Figure 7.1). The high group had generally higher average scores on the IMI, MAI, SRQ, and MS compared to the low group, indicating that they were more intrinsically motivated, more aware of their metacognitive processes, and had higher levels of self-regulation.

Figure 7.2 illustrates the memberships in the two groups at different points in time, and how one participant (`P016_AUSTIN`) changed group membership at the end of the semester. 12 participants started off the semester (`QSNR1`) in the high group, while 4 in the low group. The group membership remained the same in the middle of the semester (`QSNR2`). At semester end, 4 participants from the high group and 2 participant from the low group dropped off. One participant transitioned from high group to low group. This resulted in 7 participants in the high group and 3 participant in the low group, with no data for 6 participants at the semester end timepoint (`QSNR3`)

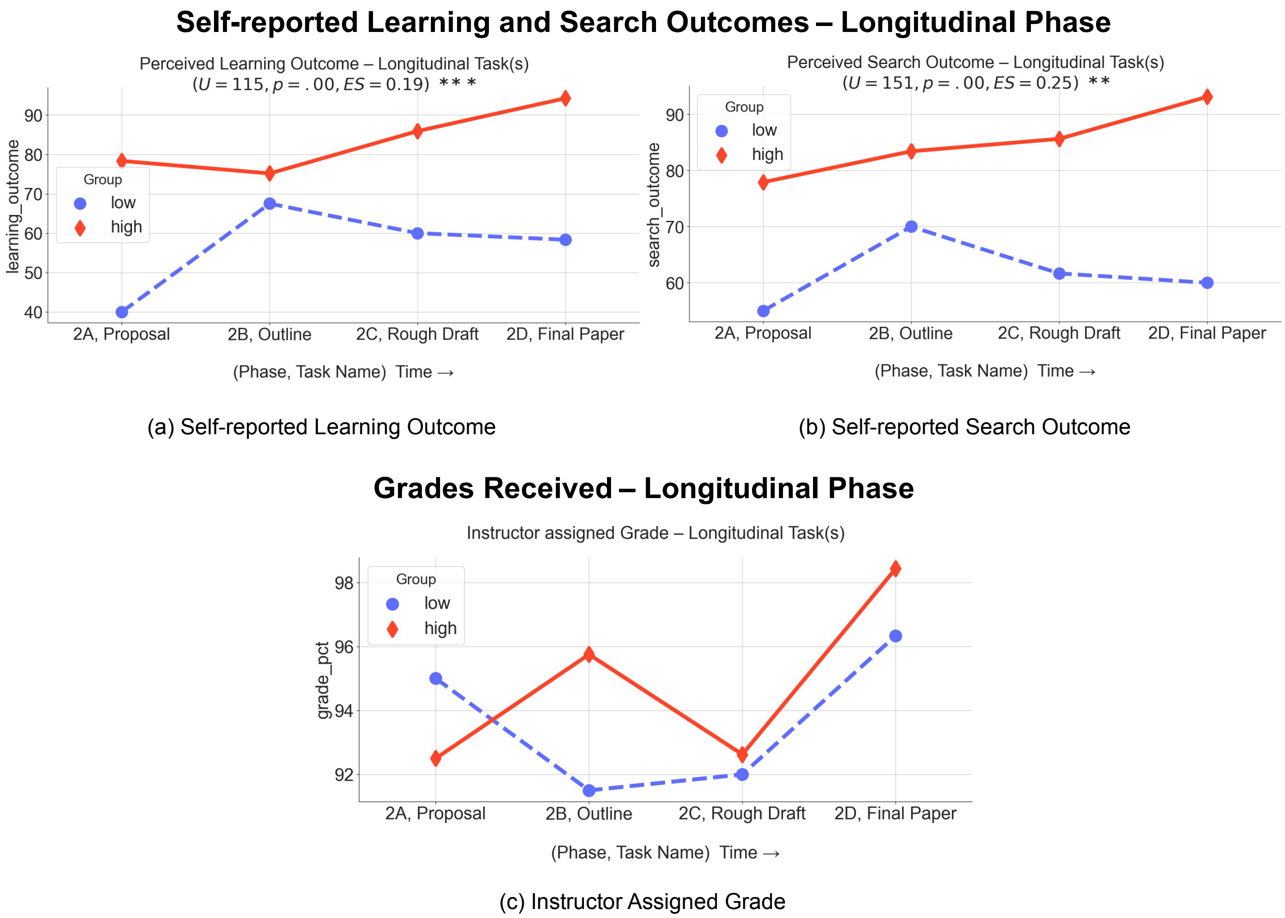


Figure 7.3: Self-reported learning and search outcomes (a, b), and instructor assigned grades for the high and low groups (c).

Henceforth, in the discussion, all the **Effect Sizes (ES)** reported as part of Mann Whitney U test compare the scores of the low group (first distribution) with those of the high group (second distribution). In other words, an example effect size $ES = 0.19$ means that there is a 19% chance that a score from the low group will be greater than the corresponding score from the high group.

7.2 Learning and Search Outcomes

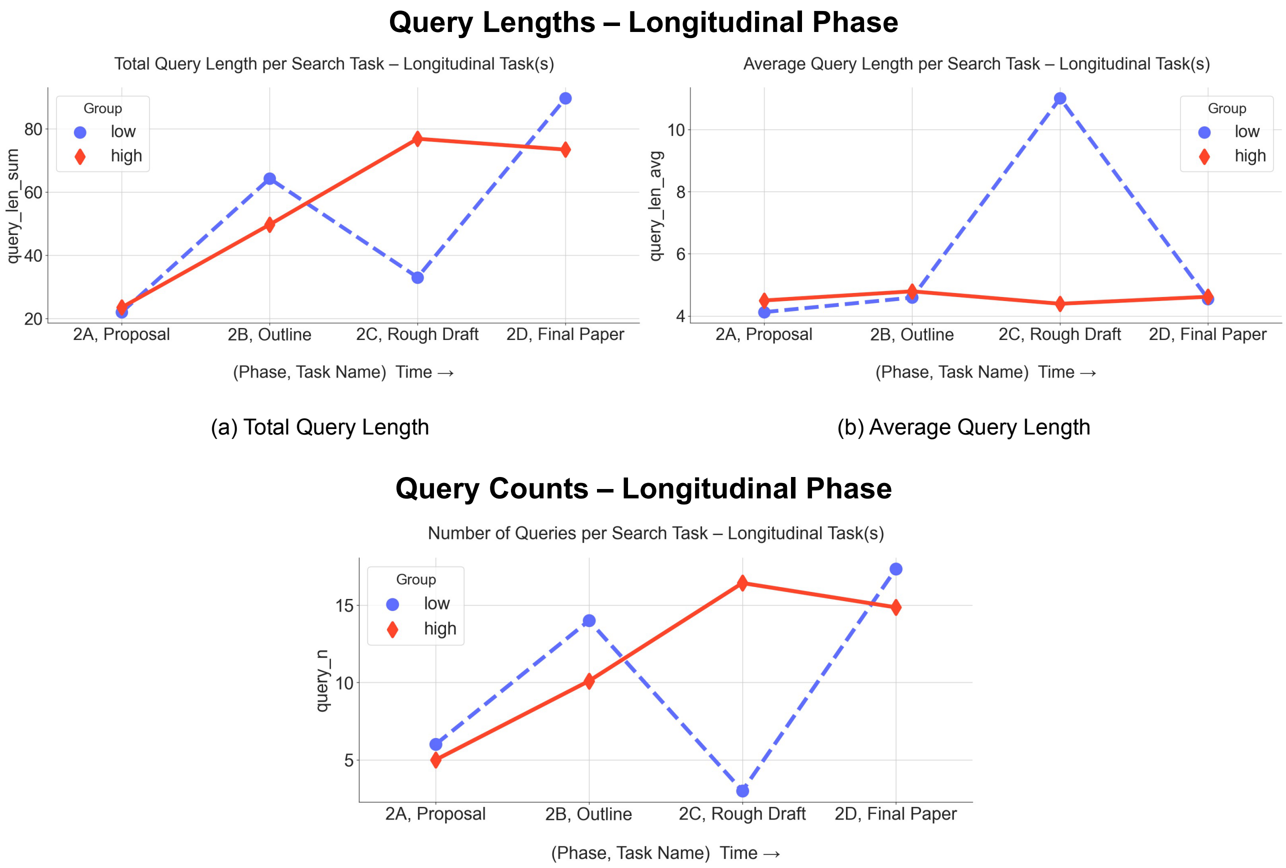
Figure 7.3 shows the mean values of the self-reported (perceived) learning outcomes (a), self-reported search outcome (b), and instructor assigned grades (c) for the Ethical Dilemma research paper writing task over the semester. The self-reported learning and search outcomes are inspired from work by Collins-Thompson et al. (2016).

We see that the high group had higher levels of perceived learning outcome and perceived search outcome compared to the low group, and these differences were statistically significant: ($U = 115.0, p = .0005, ES = 0.19$) for the learning outcome, and ($U = 151.0, p = .005, ES = 0.25$) for the search outcome. (The effect sizes indicate the probability that a value chosen at random from the low group's scores will be greater than a value chosen at random from the high group's scroes.)

The fact that the high group had statistically significant higher self-reported learning outcomes and self-reported search outcomes suggests that these students may have had a higher level of motivation, self-regulation, and better time management skills than the low group. This is because students who are motivated and self-regulated tend to be more efficient and effective in their information searching behaviours, which in turn may lead to better learning outcomes. Additionally, the high group may have been better at managing their time and resources, which would allow them to engage in more thorough and comprehensive information searching activities.

The instructor assigned grades for the high group also generally stayed higher than the low group (except for the Proposal stage). However, the differences were small. This may indicate that the instructors' grading criteria may not have fully captured the impact of information searching behaviours on learning outcomes. This is because the instructor's grading criteria may have been focused more on the content of the research paper, rather than the process of information searching. As a result, students who were more effective in their information searching behaviours may not have received a higher grade, even if their research paper was of higher quality. Another possibility is that the grading process may have been liberal towards the students.

Collins-Thompson et al. (2016) reported that "searchers' perceived learning outcomes closely matched their actual learning outcomes" and this was also indirectly correlated with their information search behaviours in terms of dwell time on documents. Let us examine in



(a) Total Query Length (b) Average Query Length

(c) Number of Queries

Figure 7.4: Lengths and count of queries for each search task in the Longitudinal Phase.

the following sections how the findings from the LongSAL study compare and contrast with those reported by Collins-Thompson et al. (2016) and others.

7.3 Q: Query Formulations

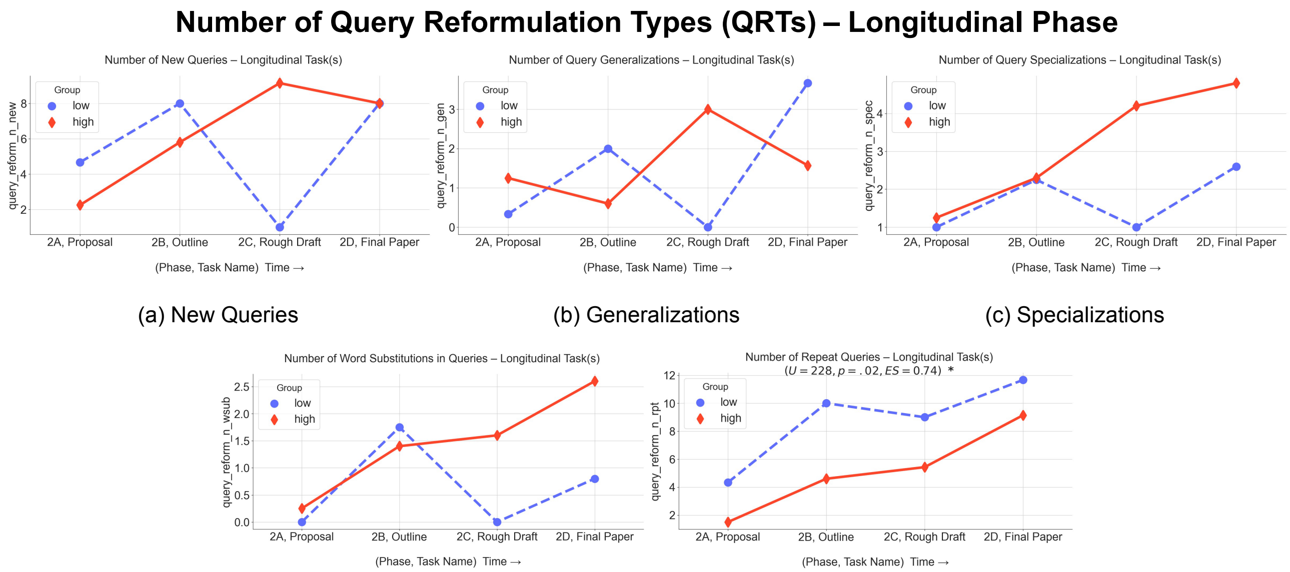
7.3.1 Length and Count of Queries per Search Task

Query length was operationalized as the number of terms (words separated by spaces) in the search query that participants submitted to the search engines or other information retrieval sites. Query length can vary from a single word to several phrases or a full sentence. Longer search queries may indicate a more specific or complex information need, while shorter search queries may be more general or broad in scope. **Queries count** per search task refers to the number of separate queries or search attempts that a participant issued in order to complete a task. This measure may vary depending on the complexity of the information need, the user's level of expertise with the search system, and other factors.

Figure 7.4 (a) and (b) shows the differences in total and average query length of the high and low groups, while Figure 7.4 (c) illustrates the number of queries issued, at different stages of writing the research paper. The low group demonstrated a zig-zag pattern in their **total query length**, and **query count**, over the duration of the semester, with a low start at proposal, followed by a peak at outline, a dip at rough draft, and again a peak at Final Paper. The high group had a steady increase in total query length and query count, from proposal to outline to rough draft, and took a very gentle dip (or remained steady) at final paper stage. Comparing the total query length and query count to the **average query length** (Figure 7.4 (b)), we see that the high group maintained a steady 4-5 terms per query throughout the semester, whereas the low group had a jump to more than 10 terms per query in the rough draft stage.

The low group issued a small number of short queries in the proposal, then a higher number of short queries during the outline, followed by an even smaller number of very long queries during the rough draft, and then a very large number of short queries during the final paper phase. On the other hand, the high group kept issuing an increasing number of similar-length (short) queries throughout the semester.

Combining these results we can posit that the low group demonstrated signs of struggling throughout the semester (Hassan et al., 2014). These students may have struggled to effectively search for information at the beginning of the semester, but then increased their search efforts as the deadlines approached. The fact that they issued a smaller number of very long queries during the rough draft phase may indicate that they were not able to effectively refine their search strategies to find more relevant and useful information. In contrast, the high group's pattern of issuing an increasing number of similar-length (short) queries throughout the semester suggests that these students may have had a more consistent and effective search strategy. They may have been better able to refine their search strategies over time, which allowed



(a) New Queries (b) Generalizations (c) Specializations

(d) Word Substitutions (e) Repeat Queries

Figure 7.5: Number of different query reformulation types (QRTs) as per taxonomy proposed by C. Liu et al. (2010).

them to find more relevant and useful information throughout the different stages of the research paper writing process.

7.3.2 Query Reformulation Types (QRTs)

Query reformulation refers to the process of modifying or refining a search query in order to improve the relevance of search results and better match the user's information needs (Section 3.2.1). Query reformulation typically occurs due to a searcher's improved understanding of how to better translate their information need into a search query. Using the taxonomy proposed by C. Liu et al. (2010) (Figure 3.3), we classified each previous-next query pair issued by participants into one of the five query reformulation types (QRTs): New, Generalization, Specialization, Word Substitution, and Repeat.

We faced a challenge in disentangling **Repeat** queries from "hub-and-spoke" behaviour, where the user goes back and forth between a SERP and different content page by using the browser's forward and back buttons [1]. Each back button press on the browser (to go back to the SERP from a content page) meant a fresh HTTP GET request was sent to the search engine. This resulted in YASBIL logging the move as a resubmission of the query. So for the discussions that follows, **"Repeat" refers to repeat queries combined with hub-and-spoke behaviour.**.

The counts of the five QRTs are presented in Figures 7.5 (a) through (e). For the low group, the trend of counts followed similarly from their trends of query counts and total query lengths (Figures 7.4 (a) and (c)), with varying intensities: alternating between high and low values at successive points in the semester. For the high group, except Query Generalizations – which followed a zig-zag pattern – all the other QRTs showed an overall increase in count throughout the semester.

The high group issued the most number of new queries and generalized queries while writing the rough draft, whereas they had the highest number of specializations and word substitutions while writing the final paper. The low group, on the other hand, had their lowest number of all types of QRTs, except repeat, while writing the rough draft. The most interesting is the trend of Query Generalizations (Figure 7.5(b)), where the high group and low group demonstrated diametrically opposite behaviour: maxima at outline and final paper for the low group, whereas minima at those stages for the high group. The high group also had significantly less number of repeat queries (or hub and spoke behaviour) throughout the semester, compared to the low group ($U = 228.0, p = .02, ES = 0.74$).

The low group's fewer counts of all QRTs while writing the rough draft suggests that they may have struggled to effectively reformulate their queries throughout the different stages of the research paper writing process, perhaps due to the complexity and depth of the research required for the tasks. The low group may have had more difficulty refining and targeting their search queries, resulting in more new and repeated queries at the final stage of the paper

[1] The SERP is the hub, which represents the initial point of inquiry, while the spokes represent the subsequent branches of exploration along different content pages. This search behaviour is often used when users have a general idea of the topic they are interested in, but need to explore different facets of the topic to narrow down their search and find relevant information.

writing process. They may also have had more difficulty with the conceptualization of their research question or topic, leading to more generalizations and fewer specializations in their queries. Additionally, their higher number of repeat queries (or hub and spoke behaviour) may indicate that they were relying on a limited set of sources or search terms, which may have limited their ability to find new and relevant information.

The high group, however, had a different pattern of query reformulation compared to the low group. They had their highest counts of new queries and query generalizations in the Rough Draft Phase, and most specialization, word substitutions, and repeat queries while writing the Final Paper. This indicates that their queries were more exploratory in the early part of the semester, and became more precise and refined in the later parts of the semester. They may have been more proactive in identifying new avenues for research earlier in the semester. The highest count of query generalizations during the writing of the rough draft may suggest that they were better able to synthesize and generalize information from their sources at an earlier stage in the writing process. The high group may also have been better able to refine their search queries through word substitutions, which peaked while writing the final paper, indicating a greater level of precision and focus in their information seeking behaviour. In contrast, the low group had their highest count of repeat queries during the outline and final paper, indicating that they may have had more difficulty finding and retaining relevant information throughout the research process.

We posit that the high group may have been more effective in their query reformulation strategies. Specifically, the high group may have been better able to identify new information needs as they worked on the rough draft, and then refine and specialize their queries as they worked on the final paper. This ability to adapt and refine their queries may have allowed them to find more relevant and useful information, which in turn may have contributed to their higher self-perceived learning and search outcomes.

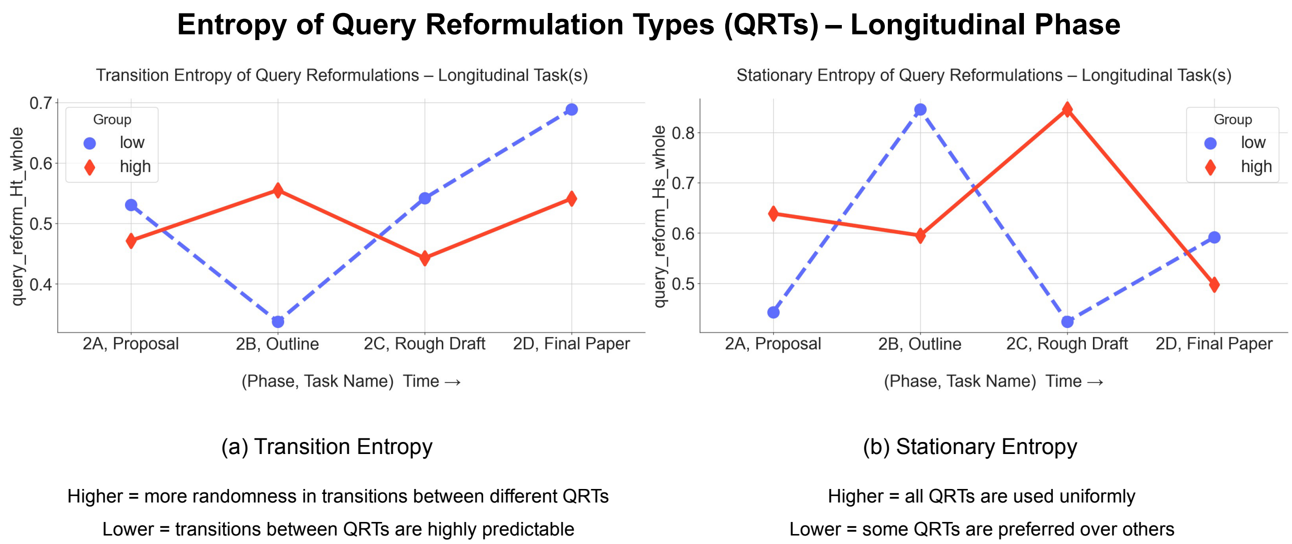


Figure 7.6: Stationary and transition entropies of query reformulation type (QRT) sequences.

7.3.3 Entropy of Query Reformulation Types

As introduced in Section 6.4, we carried out entropy analysis of query reformulation sequences produced by the participants. In the context of query reformulations, the maximum **transition entropy**, ($s \log s$), can be reached when there is an equal probability of switching between each of the $s = 5$ states, or QRTs (query reformulation types). The minimum transition entropy (0) is achieved in a fully deterministic Markov chain, where all transition probabilities are either 1 or 0. This means that with a higher transition entropy there is more randomness in the participant's transitions between different QRTs. This randomness is an indication that the participants do not have a clear progression from one QRT to another. On the other hand, a lower transition entropy indicates that the participant's transition between QRTs are highly predictable. **Stationary entropy** is calculated from the distribution of QRTs. A higher stationary entropy value indicates that the QRTs were used uniformly, while a lower stationary entropy indicates that some QRTs are preferred over others. Values of stationary entropy vary between 0 and ($\log s$), where s is the number of possible states (QRTs)[2]. All

[2] This explanation is adapted from He et al. (2016)

the entropy values presented in this chapter were normalized by their theoretical maximums, for equivalent comparison across different tasks.

The transition entropy of query reformulations followed interesting patterns for the low and high groups (Figure 7.6(a)). The low group had a V-shaped pattern showing a decrease in transition entropy from proposal to outline, then an increase from outline to rough draft to final paper. On the other hand, the high group had a zig-zag pattern, with low transition entropy during the proposal, an increase during the outline, a decrease during the rough draft, and then another increase during the final paper. This indicates that the low group had the least randomness in query reformulation strategies during the outline, while the high group had the most randomness at this stage. Subsequently, the randomness in the high group decreased, whereas that in low group increased. This suggests that during the outline stage, the low group had a more structured approach to query reformulations, compared to the high group. However, as the semester progressed, the low group's approach became more random, while the high group's approach became structured. These pattern suggests that the low group may have struggled to adapt their query reformulation strategies as they moved through the different stages of writing the paper, while the high group was more able to adjust their strategies and maintain a predictable structure in their approach.

The stationary entropy of query reformulations of the low and high groups varied over the different phases of the research paper writing process as well (Figure 7.6(b)). The low group's stationary entropy reached its maximum value at the outline phase, whereas that for the high group became maximum at the rough draft phase. This indicates that these two groups had distinct information searching behaviours throughout the writing process. The low group tried out all possible types of query reformulations during the outline (as we saw from the query reformulation counts), and then settled on using repeat queries (or hub and spoke behaviour) more, during the rough draft phase. This lowered their stationary entropy at the rough draft phase, and may have limited their ability to find new and relevant information during the later stages of the writing process.

In contrast, the high group employed all the types query reformulations with equal probability during rough draft phase (Figure **??**), resulting in a higher stationary entropy value. The increase in the high group's stationary entropy from outline to rough suggests that they were exploring a broader range of topics and concepts at this stage of the writing process, which may have allowed them to identify more relevant and useful information. The subsequent decrease in stationary entropy from rough draft to final paper suggests that they were able to narrow down their focus and consolidate their understanding of the subject-matter as they progressed, which may have contributed to their higher self-perceived learning and search outcomes.

7.4 L: Interaction with Search Results / Source Selection / Item Selection

7.4.1 Number of Clicks per Query

ResN6629066624252186314 | ResP5466996162614873067 | ResP1250472196403953856

Number of clicks per query refers to the number of times a participant clicked a link on a search result page after conducting a search query. This metric reflects the level of interaction and engagement of the participant with the search results, as well as their ability to assess the relevance and usefulness of each search result presented to them. A higher number of clicks per query may indicate that the participant is more engaged and willing to explore a wider range of information sources, while a lower number may suggest a more focused and targeted search approach.

The high group had fewer total clicks per query (ResN6629066624252186314), and a lower average number of clicks per query (ResP5466996162614873067), compared to the low group, across most of the tasks during the semester. The low group had the highest average number of clicks per query when writing the final paper, while the high group had the highest average number of clicks per query earlier in the semester, during the paper outline phase. Additionally, the high group had lower variability (standard deviation) in the number of clicks per query compared to the low group ($U = 1235.0, p = .00, ES = 0.68$).

This suggests that the high group was able to achieve better search outcomes with fewer clicks, indicating a higher level of efficiency in their search strategies. They may have been able to refine their search strategies and become more efficient as the semester progressed, resulting in fewer clicks per query when writing the final paper. This indicates the high group's ability to better assess the relevance and usefulness of search results, and to use their knowledge and cognitive strategies more effectively during the search process. In contrast, the low group may have struggled to improve their search strategies, resulting in a higher number of clicks per query when writing the final paper.

It is also possible that the differences in average clicks per query between the two groups during different phases of the semester reflect differences in the complexity or specificity of the information needed for the different tasks. For example, the paper outline phase may have required more broad and exploratory searches, while the final paper phase may have required more targeted and specific searches.

7.4.2 Dwell Time on Search Results

7.4.2.1 Publication Search Results

ResP8346620135317238453 | ResN4762872426347949231

Contrary to dwell time on Web SERPs, the low group had longer average dwell time on academic publication search results, in all the four stages, compared to the high group (ResN4762872426347949231) ($U = 149.0, p = .05, ES = 0.72$). Total dwell time of the low group was also longer, except at the final paper stage, when the high group surpassed the low group in dwelling on publication search results.

This suggests that the low group spent more time examining and considering the academic publication search results compared to the high group. It is possible that the low group's lower levels of motivation, self-regulation, and metacognition may have led them to spend more time on search results as they may have found it harder to quickly evaluate and assess the relevance of search results to their task. On the other hand, the high group's higher levels of motivation, self-regulation, and metacognition may have enabled them to quickly identify relevant search results and move on to the next stage of their task. However, at the final paper stage, the high group spent more time on search results as they may have needed to ensure that they had not missed any important information and had thoroughly covered their topic.

7.4.2.2 Web Search Results (SERPs)

ResP2704210890287754729 | ResN2852099729940598719

We examined how dwell time changed across the stages of the search process for each group. For the high group, dwell time decreased from proposal to outline, then increased from outline to rough draft, and then decreased again during the final paper stage. This suggests that participants in the high group spent less time on SERPs during the outline stage but spent more time during the rough draft stage. However, they spent less time again during the final paper stage. For the low group, dwell time decreased from proposal to outline and then reached a minimum during the rough draft stage. However, it increased again during the final paper stage. This suggests that participants in the low group spent the least amount of time on SERPs during the outline and rough draft stages but spent more time during the final paper stage.

This indicates at the positive relationship between individual differences in motivation, self-regulation, and metacognition and the amount of time spent on SERPs during a search task.

The high group's higher levels of motivation, self-regulation, and metacognition may have enabled them to engage in more effective and efficient search behaviours, leading to longer dwell times on SERPs. The higher dwell time on SERPs for the high group may indicate that they are more strategic and deliberate in their search behaviour, taking the time to carefully examine search results and evaluate their relevance to their information needs. This could be related to their higher levels of motivation, self-regulation, and metacognition, which enable them to set goals, monitor their progress, and adjust their strategies as needed to achieve optimal outcomes.

In contrast, the low group's lower levels of these individual differences may have hindered their ability to engage in effective search behaviours, leading to shorter dwell times on SERPs. This suggests a more haphazard approach to search behaviour, with less attention to evaluating search results and less ability to regulate their search strategies. This may be related to their lower levels of motivation, self-regulation, and metacognition, which may limit their ability to set clear goals, monitor their progress, and make effective decisions in their search behaviour.

In general, the high group had longer dwell times on web search results, and shorter dwell times on publication search results. The low group demonstrated the opposite pattern. This is an interesting finding. It suggests that the high group may be more efficient in quickly finding and evaluating information on the web, while the low group may take longer to navigate and process the same information. On the other hand, the low group may be more thorough in searching and evaluating academic publications (which was arguably one of the main aspects of writing the research paper), while the high group may be able to quickly identify and select relevant publications.

7.5 I: Interaction with Sources / Content Pages / Information Objects

7.5.1 Academic Publications

The low group viewed more number of academic publications earlier in the semester, while writing the outline, while high group viewed more number of publications later in the semester, while writing the rough draft and the final paper. This indicates that the low group had a more exploratory approach towards their research earlier in the semester, as they spent more time looking through a larger number of academic publications while writing the outline. On the other hand, the high group had a more focused approach, spending more time on fewer publications during the later stages when writing the rough draft and the final paper. This may be an indication of a difference in information-seeking behaviour and strategies between the two groups, with the low group possibly using a more broad and inclusive approach to gather information, while the high group uses a more selective approach. It could also reflect differences in the stage of research and writing that the two groups were at during each stage, with the low group at a more exploratory stage earlier in the semester, while the high group was at a more advanced stage of writing.

The total and average dwell time (possibly reading time) also followed a similar pattern. The high group had longer dwell time on publications in total and on average, in the later parts of the semester, while the low group dwelt more on publications in the earlier parts of the semester. This suggests that the high group had a more focused and deliberate approach to their research process, possibly due to better metacognitive and self-regulation skills. They spent more time on each publication, reading and analysing it thoroughly, in the later parts of the semester when they were working on the rough draft and final paper. This behaviour could have led to a deeper understanding of the content and better integration of the information into their writing. On the other hand, the low group may have had a more scattered approach, possibly due to lower levels of motivation, self-regulation, and metacognitive awareness, resulting in spending less time on each publication and viewing a larger number of publications earlier in the semester. This behaviour could have resulted in a superficial understanding of the content and difficulties in integrating the information into their writing.

7.5.2 Non-scholarly / Normal Content pages

In a similar vein to publication vs web search results, the low group visited more number of content pages that were not academic publications compared to the high group. Their total dwell time on such pages were also longer than the high group, but the average dwell time was shorter. This suggests that the low group engaged in more exploratory behaviour, perhaps looking for additional sources of information or inspiration, beyond the academic publications. However, their lower average dwell time on these pages indicates that they may have been quickly scanning or skimming through the content, rather than engaging with it in depth. It's possible that they were spending less time on average on these pages because they were not as relevant or useful for their research paper.

On the other hand, the high group may have been more focused on finding academic sources for their paper, which could explain their lower number of visits to non-academic content pages. They may have been more focused on the task at hand and less likely to be distracted by extraneous information. The longer average dwell time on content pages suggests that they were more deeply engaging with the material they did visit. Overall, the low group may have been more inclined to explore a broader range of information sources, while the high group was more focused on academic sources and more likely to engage deeply with the content they did view.

7.6 Entropy of Search Tactic Sequences

Similar to entropy analysis of Query Reformulation types (Section **??**), we perform entropy analysis of search tactic transitions. This analysis is directly inspired from He et al. (2016), who in turn were inspired from Krejtz et al. (2015).

In the context of search tactic transitions, a higher value of transition entropy (H_t) indicates more randomness and uncertainty in the participant's search behaviour (which is composed of different search tactics, and transitioning or switching between those tactics). A lower value of transition entropy indicates that the search behaviour (i.e. tactic switching behaviour) is highly predictable. For stationary entropy of search tactics (H_s), a higher value indicates that participants utilize all the search tactics with equal probability, while a lower value suggests that certain search tactics are favoured over others. The entropies discussed in the following sections were normalized by their theoretical maximums, for equivalent comparison across different conditions.

ResN7289045485214067875 Both groups had highest Ht during 2C (and also 3 Finance?)

ResP5638176378435107439 - beginning ResN3435548740759333098 - middle ResP6845508037293675246 - end

No sig diff for whole task, and at the beginning of task Sig diff appeared at the middle and increased at the end of task (similar to as observed by Gwizdka).

ResP8685879301269737254 LO group had hghest Hs in 2C No sig diff in overall task

No diff at beginning. Diff is most prominent at middle (ResN1140102300744288543) and also somewhat at end (ResN663758335140886910)

7.7 Search Result Pages vs Content Pages

Overall, the high group visited more search result pages and content pages as the semester progressed, while the low group had a drop in these behaviours after the outline phase. The difference in visits to Search Results pages was almost significant ($U = 133.5, p = .06, ES = 0.33$). This suggests that the high group may have been more engaged and persistent in their information searching behaviours, while the low group may have experienced a decrease in motivation or self-regulation as the semester progressed, specifically before writing the rough draft.

It is also noteworthy that the low group had a rebound in the number of pages visited while writing the final paper. This could indicate that they were able to re-engage with the task and their information searching behaviours improved as the final-paper deadline approached.

The fact that the total dwell time on webpages followed a similar pattern as the number of pages visited is also interesting. It suggests that the high group spent more time engaging with the content they found, which could have contributed to their better learning outcomes.

8

Results: Repeated vs New Search Tasks

Same LPA profiles from Section ??.

8.1 Learning Outcomes

8.2 Querying Behaviour

8.2.1 Length of Queries

8.2.2 Number of Queries per Search Task

8.2.3 Number of Clicks per Query

8.2.4 Query Reformulation Types (QRTs)

8.3 Entropy Analysis of Querying Behaviour

9
Discussion

Quick Check Results (Descriptive)

- which results have the strongest effect sizes and RBC scores?
- which users went beyond page 1 in a SERP?

Hypotheses from He et al. (2016):

The second set (H2) compares two different user groups, experts and novices, using one of the search systems in two different conditions. The H2 hypotheses illustrate how a focus on search tactics provides a different lens to view search logs.

- H2.1: Search experts are likely to be more predictable in their choice of search tactics compared to novices
- H2.2: Search experts have developed a set of search tactics they prefer over others, while novices use search tactics more uniformly.
- H2.3: While working with a search system novices will find a preferred method of transitioning from one search tactic to another. In other words, their search tactics transitions will become more predictable over time.

- H2.4: While working with a search systems novices will find preferred search tactics to use. In other words, their distribution of search tactics will become less uniform over time.

9.1 Summary of Results

Number of clicks per query:

- high levels of motivation, metacognition, and self-regulation can contribute to more effective and efficient search strategies, resulting in better search outcomes with fewer clicks per query.
- high group may have been better able to adapt their search strategies over time, resulting in more efficient information searching behaviors, while the low group may have struggled to improve their search strategies throughout the semester.
- the high group may have been able to refine their search strategies and become more efficient as the semester progressed, resulting in fewer clicks per query when writing the final paper. In contrast, the low group may have struggled to improve their search strategies, resulting in a higher number of clicks per query when writing the final paper.
- It is also possible that the differences in average clicks per query between the two groups during different phases of the semester reflect differences in the complexity or specificity of the information needed for the different tasks. For example, the paper outline phase may have required more broad and exploratory searches, while the final paper phase may have required more targeted and specific searches.

Query Reformulation Counts:

- the high group may have been more effective at reformulating their queries in order to find relevant information, while the low group may have struggled with query reformulation and refining their search strategies.

- Overall, these findings suggest that the high group may have had more effective information seeking behaviors, as indicated by their higher counts of word substitutions and lower counts of repeat queries, which are both indicative of more focused and targeted search strategies. The low group, on the other hand, may have struggled with information overload and refining their search queries, which is reflected in their higher counts of new queries, generalizations, specializations, and repeat queries.

SERP Dwell Time

- Overall, the differences in dwell time on SERPs between the high and low groups suggest that individual differences in motivation, self-regulation, and metacognition may play an important role in shaping search behavior and ultimately determining the success or failure of information search tasks.

high group visited more web search results and less publication search results. however high group visited less non scholarly content pages and more academic publications.

Struggling vs. Exploring

Indicators predictive of struggling (Hassan et al., 2014):

- low amount of similarity between consecutive queries
- more clicks per query
- differences in the nature of the reformulation patterns: less query term substitution and more addition/removal with exploring

9.2 RQ1: how do search behaviours change over time?

- Phase 1
  - Query reform measures
  - search measures
  - learning outcomes

  - SHEG tasks

- Phase 2

- Phase 3

- Indicators predictive of struggling (Hassan et al., 2014):

  - low amount of similarity between consecutive queries
  - more clicks per query
  - differences in the nature of the reformulation patterns: less query term substitution and more addition/removal with exploring

9.3 RWx: Entropy based research questions from He 2016

9.4 RQ2: similarities and differences in repeated vs new tasks

Phase 1 and Phase 3 only

- search task
- SHEG tasks
- Similarities -> measures having no sig diff

  - Overall task, Beginning, middle, end

- Differences -> measures having sig diff

  - Overall task, Beginning, middle, end

SES3 SHEG - only one participant issued a query; others did not

9.5 RQ3: correlation with learning?

10 Discussion and Conclusion

see Jacek's thesis

10.1 Research Summary

By using Latent Profile Analysis to group participants based on their self-reported scores on motivation, metacognition, and self-regulation, we were able to identify distinct subgroups within the sample. This can help us to better understand how individual differences in these factors might relate to information searching behaviors and learning outcomes.

Using a longitudinal design with with stages of writing (proposal, outline, rough draft, and final paper) also allows us to examine how these factors and behaviors might change over time and how they might relate to each other at different stages of the writing process.

Overall, the study has the potential to provide valuable insights into the complex interplay between individual differences, information searching behaviors, and learning outcomes in the context of research paper writing.

QRTs Overall, these findings suggest that effective query reformulation strategies may be an important factor in achieving better self-perceived learning and search outcomes among undergraduate students writing research papers. Future research could explore the specific techniques and strategies that are most effective in promoting effective query reformulation among undergraduate students.

10.2 Summary of Results

which results have the strongest effect sizes and RBC scores?

CONCLUSION: obtained grades are not good indicators of learning

10.3 Methodology

10.4 Contributions

- Latent Profile Analysis and Latent Profile Transition Analysis
- Entropy
- Longitudinal Study Design
- Combining Motivation Metacognition Self regulation

10.4.1 Latent Profile Analysis

Additionally, LPA can identify subgroups of individuals who may be at risk for negative outcomes, such as poor task performance, low job satisfaction, or developmental delays. This information can be used to design targeted interventions that are tailored to the specific needs of these subgroups.

In summary, LPA is a powerful and flexible person-centered statistical approach that is increasingly used in organizational psychology and child development research. Its ability to identify and describe latent subgroups within a population provides a more nuanced understanding of human behavior, and its applications have the potential to inform the design of targeted interventions that can improve outcomes for individuals and organizations.

In the context of information search behavior, LPA can be used to investigate how search behavior varies across different contexts or populations. For example, researchers may use LPA to identify different search patterns among novice versus expert users, or among different age or cultural groups. This can help to inform the design of search interfaces and other information systems that are tailored to the needs of specific user groups.

Overall, LPA is an interesting and valuable technique for understanding information search behavior because it allows researchers to move beyond simple group comparisons and identify meaningful subgroups within a population. By uncovering these subgroups, researchers can gain deeper insights into the complex and nuanced ways that people search for information online, and use this knowledge to improve the design and effectiveness of information systems.

Overall, the use of LPA to classify participants into high and low groups based on their scores on the IMI, MAI, and SRQ was an effective method for identifying individual differences in motivation, metacognition, and self-regulation. This allowed us to examine the relationship between these individual differences and search behavior, which is the focus of this study.

10.4.2 Entropy

Transition analysis and entropy are powerful tools in the field of information retrieval that can help to uncover patterns in user search behaviour across disparate tasks and activities. By analysing transitions between different states, such as between queries, documents, or web pages, transition analysis can reveal important insights into how users navigate information spaces and interact with search systems. For example, it can be used to identify common search patterns or to identify areas where users may be experiencing difficulty or confusion.

Entropy, on the other hand, provides a quantitative measure of the uncertainty or randomness in user search behaviour. By calculating entropy values for different aspects of search behaviour, such as query reformulation or click-through behaviour, researchers can gain a better understanding of the structure and characteristics of user search behaviour in different contexts. This can help to identify areas for improvement in search systems and to design more effective search algorithms and interfaces.

Together, transition analysis and entropy provide a powerful framework for understanding user search behaviour and improving the design and effectiveness of search systems. By analysing transitions and quantifying the level of uncertainty in user behaviour, researchers can uncover valuable insights that can inform the development of more effective search systems and improve the overall user experience.

From ASIST award session

FOLLOW UP WITH ...

- Change of the self, self-reflection
- Look at anthropology perspective

ROB:

- Look at qualitative data. More interesting

HEATHER:

- what can we share back to the teachers of the course, librarians and others

10.5 Limitations

10.5.1 Theoretical Limitations

Learning outcomes Cite Urgo & Arguello (2022)

10.5.2 Technical Limitation

- No PDF
- N=16 to N=10
- Also check anticipated limitations section from proposal
- Did not assess goodness of fit of whether first order Markov chain is the most appropriate for modelling search tactic transitions (Besag and Mondal's (2013) statistical test)
- did not track revisits across different tasks

- clicks may have been on popup ads, cookie consent forms, and others, instead of only links (difficult to guesstimate from HTML markup)

Some difficulties:

- opening same URL in 2 different tabs and switching between them for a while before realising one is different from the other
  - or maybe long webpage (e.g.) book - two tabs at two different locations in webpage

10.6 Future Work

We have barely scratched the surface

- What factors are specifically responsible for the transitions or the changes in the Latent profiles at different points in the semester?
- Correlation between user profiles and search tactics (Table 9 Taramigkou et al., 2018)
- Parallel Browsing Behaviour
- Consistency of search terms (from FIRE group) – low consistency = evidence of learning?
- Navigators vs Explorers (from IWSS)
- Struggling vs exploring
- Strategies from (Lam et al., 2007) section 6: short nav, topic explore, etc
- Bing + ChatGPT, from MSR blogpost:
  - `https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web/`
- Google + ChatGPT, from Sundar Pichai
  - `https://blog.google/technology/ai/bard-google-ai-search-updates/`
  - “AI features in Search can distill information to help you see the big picture.”

Contrary opinions:

- https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web
  - "When we're dealing with sequences of words, lossy compression looks smarter than lossless compression."
  - hallucinations
- https://abcnews4.com/news/nation-world/experts-warn-dangers-ai-google-microsoft-push-new-chatbots-chatgpt-bard-technology-big-tech-open-ended-questions-responses-alphabet-misinformation-disinformation-internet-new-developments-bing-search-engines
  - Matt is here
  - potential to misinformation

  > *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/

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

10.7 New references

Be sure to add new references

- https://www.mdpi.com/2673-4001/4/1/8
- https://link.springer.com/article/10.1007/s10648-023-09730-8
- Claudia Hauff's works from CHIIR2020 onwards
- SYR and other's works

Appendices

A
Prior Work: Pilot Study

A.1 SES1: Initial Session

B

`QSNR`: Questionnaires

B.1 `QSNR0`: Recruitment Questionnaire

Thank you SO much for your willingness to participate in the LongSAL research study. The aim of this study is to identify how search engines can better support the needs of university students' learning and education. To be eligible for this study, you must be enrolled in the I 303 Ethical Foundations for Informatics course for the Spring 2022 semester. Please fill out the information requested below. We will get back to you if you are selected to participate in the study. The principal investigator, Nilavra Bhattacharya, can be reached at `<email-address>` for any questions or concerns.

1. Please select which section of the I-303 Ethical Foundations for Informatics course you are enrolled in.

    - TUE: FLEISCHMANN, VERMA
    - WED: FLEISCHMANN, GURSOY
    - THU: FLEISCHMANN, BAUTISTA
    - FRI: FLEISCHMANN, DAY

2. Please select the degree level/name of the program you are in.

    - Bachelor's
    - Master's
    - Integrated Bachelor's and Master's
    - PhD
    - Other .....

3. Please state which year of the program you are in.

    - Freshman
    - Sophomore
    - Junior
    - Senior
    - Graduate Year 1
    - Graduate Year 2
    - Other .....

4. Please state your major(s) .....

5. Do you have native-level familiarity with English language?

    - Yes
    - No
    - Other: .....

6. Please state your age (in years) .....

7. Please state your gender .....

8. With which ethnicities do you identify? Please select all that apply:

    - African

- African American / Black
- Asian - East
- Asian - South East
- Asian - South
- Asian - Middle East
- Caucasian / White
- Hispanic / Latinx
- Native American
- Pacific Islander
- Mixed
- Other .....

9. Are you an international student? If “yes”, where are you originally from?

    - Yes .....
    - No

10. We need your contact information to communicate with you over the semester (if you are selected). Your contact information will not be used in any other way, and will be kept private. Please enter an email address that you check regularly. We will use this email address to send communications and Amazon Gift Cards as payment. .....

11. Your name as you would like us to address you (solely for communication). .....

B.2 QSNR1 - QSNR3: Entry, Mid-term and Exit Questionnaires

B.2.1 Consent Form

Consent to Participate in Research

Basic Study Information:

- UT Austin IRB Approved
- **Submission ID:** STUDY00002136
- **Date Approved:** December 8, 2021
- **Title:** LongSAL: A Longitudinal study on Searching as Learning
- **Principal Investigator:** Nilavra Bhattacharya, PhD Student, School of Information, UT Austin
- **Faculty Advisor:** Jacek Gwizdka, Associate Professor, School of Information, UT Austin

Invitation to be Part of a Research Study Things you should know:

- The purpose of the study is to identify how search engines can be improved to better support university students' learning and education.
- In order to participate, you must be enrolled in the I303 Ethical Foundations for Informatics course in the Spring 2022 semester.
- If you choose to participate, you will be asked, over the course of the semester, to take three surveys (10-15 mins each), attend two Zoom sessions (60-90 mins each), and record browsing activity while working on Final Project Paper. All parts of the study will be conducted online.
- Risks or discomforts involved in this research study are not greater than everyday life.
- There is no direct benefit for participating in this study.
- Taking part in this research study is voluntary. You do not have to participate, and you can stop at any time.

More detailed information may be described later in this form.

Please take time to read this entire form and ask questions before deciding whether to take part in this research study.

What is the study about, and why are we doing it? The aim of this longitudinal study is to identify how university students search the web for educational research activities.

Findings from this study will help to understand how search engines can be improved to better support university students' learning and education, and therefore help to build more human-centred and learning-centric search systems.

What will happen if you take part in this study? If you agree to take part in this study, you will be asked to perform the following activities over the duration of the Spring 2022 semester.

How long will this study take and how many people will be in the study? Participation in this study will take approx. 10-15 minutes each for the three surveys, and 60-90 minutes each for the two synchronous Zoom sessions. There will be about 30-40 participants in total in this study.

What risks and discomforts might you experience from being in this study? There are no major foreseeable risks to participating in this study. There may be a very minimal potential risk of confidentiality, or possibly frustrations with some tasks. To address the risk of confidentiality, once data collection is completed, all personally identifying data will be destroyed by erasing all digital files and shredding all the physical records. To address the risk of frustration, you can move at your own pace, or stop whenever you wish.

How could you benefit from this study? You will receive no direct benefit from participating in this study; however, this study will help to improve our current understanding of how students search the web for education and learning-related goals over time. This will the development of better learning-centric search systems.

What data will we collect from you? As part of this study, we will collect your:

- audio, screen recordings, and browsing logs when you are participating in the Zoom sessions. You can turn off your face video during these sessions.
- browsing log data, when you are performing research for the course final project. You can start and stop the logging when you choose.
- anonymized submissions for the final project at various points in the Semester
- self-reported grades for the final project assignments that you received

The holistic data about your research-assignment related internet search activity, the material that you produce for your assignments, and the scores you receive for those assignments, will help us understand where students perform well, where things can be improved, and how search engines can be improved to better support university students' learning and education.

How will we protect your information? We will protect the privacy and the confidentiality of your data by:

- Assigning you a coded username at the beginning of the study to protect confidentiality, and all your submitted data will be linked to this coded username.
- All digital data generated in the study will be stored via a university-approved secure cloud-based storage service and password-protected computers. All computers used in the project are password protected.
- Audio recordings will be listened to only for research purposes. Audio recordings will be transcribed and coded. No information that can be used to uniquely identify an individual will be present

We may share your data with other researchers for future research studies that may be similar to this study or maybe very different. In these cases, the data shared with other researchers will NOT include any information that can directly identify you.

We plan to publish the results of this study. To protect your privacy, we will NOT include any information that could directly identify you.

What will happen to the information we collect about you after the study is over? We will keep your research data to use for future analyses and publications. Any information that can directly identify you will be deleted from the research data collected as part of the project.

How will we compensate you for being part of the study? You will receive up to USD 150 in Amazon Gift cards if you complete all the components of the study, as described above. If you choose to withdraw early from the study, you will receive compensation for the parts you have completed. You will be responsible for any taxes assessed on the compensation.

Your Participation in this Study is Voluntary It is totally up to you to decide to be in this research study. Participating in this study is voluntary. Your decision to participate will not affect your relationship with The University of Texas at Austin. You will not lose any benefits or rights you already had if you decide not to participate. Even if you decide to be part of the study now, you may change your mind and stop at any time. You do not have to answer any questions you do not want to answer.

This Study is NOT a part of the I 303 Ethical Foundations of Informatics course:

- The study researchers intend to only recruit participants from the student pool enrolled in the course.
- The course instructors will not share any student data with the researchers.
- The course instructors will not be aware of which students did or did not participate in this study.
- Participation in this study is completely voluntary, and in no way affects the outcome of this course for you, or your academic relations with the course instructors.

Even after consenting to participate, you can choose to withdraw consent anytime during the semester. We will delete all the data collected from you up to that point. You will receive compensation for the parts you have completed, as outlined above.

Contact Information for the Study Team and Questions about the Research Prior to, during, or after your participation, you may contact the researchers below if you have any questions about this research, or feel you may have been harmed due to participation:

Nilavra Bhattacharya

Phone: `<phone-number>`

Email: `<email-address>`

Or

Jacek Gwizdka

Email: `<email-address>`

Contact Information for Questions about Your Rights as a Research Participant If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the following:

The University of Texas at Austin

Institutional Review Board

Phone: `<phone-number>`

Email: `<email-address>`

Please reference the study protocol number (STUDY00002136) in your communications.

Your Consent By clicking the button below, you are agreeing to participate in this study. Make sure you understand what the study is about before you consent. If you have any questions about the study after you consent, you can contact the study team using the information provided above.

- By clicking this button I understand what the study is about and my questions so far have been answered. I agree to participate in this study.

Please enter the coded username assigned to you (shared via email) . . . ..

B.3 Motivation

Adapted from Intrinsic Motivation Inventory (IMI) (Ryan, 1982). Items will be randomly ordered.

Scoring directions: Score each response from 1 (not at all true) to 5 (very true). Then reverse score the items marked with **(R)**. To do that, subtract the item response from 6, and use the resulting number as the item score. Then, calculate subscale scores by averaging across all the items on that subscale. The subscale scores are then used in the analyses of relevant research questions.

For each of the following statements, please indicate how true it is for you, using the following scale:

(1) not at all true — somewhat true — very true (5)

B.3.1 Interest/Enjoyment

1. I will enjoy taking this course very much.
2. This course will be fun to do.
3. I think this will be a boring course. **(R)**
4. This course will not hold my attention at all. **(R)**
5. I would describe this course as very interesting.
6. I think this course will be quite enjoyable.

B.3.2 Perceived Competence

1. I think I will be pretty good at this course.
2. I think I will be doing pretty well at this course, compared to other students.
3. After working at this course for awhile, I will feel pretty competent.
4. I think I will be satisfied with my performance in this course.
5. I think I am pretty skilled at this course.
6. This is a course that I think would not be able to do very well. **(R)**

B.3.3 Effort/Importance

1. I plan to put a lot of effort into this course.
2. I don't think I will try very hard to do well at this course. **(R)**
3. I will try very hard on this course.
4. It is important to me to do well in this course.
5. I do not plan to put much energy into this course. **(R)**

B.3.4 Value/Usefulness

1. I believe the course and the final project activities could be of some value to me.
2. I think that doing the final project activities is useful for me.
3. I think the final project is important activity to do because it can equip me with skills that are necessary for making ethical decisions in my adult and professional life.
4. I would be willing to do research on the final project topic again because it has some value to me.
5. I think doing the final project activities will help me in my adult and professional life
6. I believe doing the final project activities will be beneficial to me.
7. I think this is an important course.

B.3.5 Pressure / Tension (not in `QSNR1`)

1. I do not feel nervous while doing the final project activities. **(R)**
2. I feel very tensed while doing the final project activities.
3. I am very relaxed while doing the final project activities. **(R)**
4. I feel anxious while working on the final project parts.
5. I feel pressured while doing the final project activities.

B.3.6 Perceived Choice (not in `QSNR1`)

1. I believe I have some choice about doing the final project activities.
2. I feel like it is not my own choice to do the final project parts. **(R)**
3. I don't really have a choice about doing the final project tasks. **(R)**
4. I feel like I have to do the final project tasks. **(R)**
5. I do the final project activities because I have no choice. **(R)**
6. I do the final project activities because I want to.
7. I do the final project activities because I have to. **(R)**

B.4 Self-regulation

Self-Regulation Questionnaire (SRQ) by J. M. Brown et al. (1999).

Please answer the following questions by selecting the option that best describes how you are. There are no right or wrong answers. Work quickly and don't think too long about your answers.

(1) Strongly Disagree – Disagree – Neutral – Agree – Strongly Agree (5)

1. I usually keep track of my progress toward my goals.
2. My behavior is not that different from other people's. **(R)**
3. Others tell me that I keep on with things too long. **(R)**
4. I doubt I could change even if I wanted to. **(R)**
5. I have trouble making up my mind about things. **(R)**
6. I get easily distracted from my plans. **(R)**
7. I reward myself for progress toward my goals.
8. I don't notice the effects of my actions until it's too late. **(R)**
9. My behavior is similar to that of my friends. Evaluating
10. It's hard for me to see anything helpful about changing my ways. **(R)**
11. I am able to accomplish goals I set for myself.
12. I put off making decisions. **(R)**
13. I have so many plans that it's hard for me to focus on any one of them. **(R)**
14. I change the way I do things when I see a problem with how things are going.
15. It's hard for me to notice when I've "had enough" (alcohol, food, sweets, internet, social media) **(R)**
16. I think a lot about what other people think of me.
17. I am willing to consider other ways of doing things.
18. If I wanted to change, I am confident that I could do it.
19. When it comes to deciding about a change, I feel overwhelmed by the choices. **(R)**

20. I have trouble following through with things once I've made up my mind to do something. **(R)**
21. I don't seem to learn from my mistakes. **(R)**
22. I'm usually careful not to overdo it when working, eating, drinking, or being on social media.
23. I tend to compare myself with other people.
24. I enjoy a routine, and like things to stay the same. **(R)**
25. I have sought out advice or information about changing.
26. I can come up with lots of ways to change, but it's hard for me to decide which one to use. **(R)**
27. I can stick to a plan that's working well.
28. I usually only have to make a mistake one time in order to learn from it.
29. I don't learn well from punishment. **(R)**
30. I have personal standards, and try to live up to them.
31. I am set in my ways. **(R)**
32. As soon as I see a problem or challenge, I start looking for possible solutions.
33. I have a hard time setting goals for myself. **(R)**
34. I have a lot of willpower.
35. When I'm trying to change something, I pay a lot of attention to how I'm doing.
36. I usually judge what I'm doing by the consequences of my actions.
37. I don't care if I'm different from most people. **(R)**
38. As soon as I see things aren't going right I want to do something about it.
39. There is usually more than one way to accomplish something.
40. I have trouble making plans to help me reach my goals. **(R)**
41. I am able to resist temptation.
42. I set goals for myself and keep track of my progress.
43. Most of the time I don't pay attention to what I'm doing. **(R)**
44. I try to be like people around me.

45. I tend to keep doing the same thing, even when it doesn't work. **(R)**
46. I can usually find several different possibilities when I want to change something.
47. Once I have a goal, I can usually plan how to reach it.
48. I have rules that I stick by no matter what.
49. If I make a resolution to change something, I pay a lot of attention to how I'm doing.
50. Often I don't notice what I'm doing until someone calls it to my attention. **(R)**
51. I think a lot about how I'm doing.
52. Usually I see the need to change before others do.
53. I'm good at finding different ways to get what I want.
54. I usually think before I act.
55. Little problems or distractions throw me off course. **(R)**
56. I feel bad when I don't meet my goals.
57. I learn from my mistakes.
58. I know how I want to be.
59. It bothers me when things aren't the way I want them.
60. I call in others for help when I need it.
61. Before making a decision, I consider what is likely to happen if I do one thing or another.
62. I give up quickly. **(R)**
63. I usually decide to change and hope for the best. **(R)**

Scoring Directions: Score each response from 1 (strongly disagree) to 5 (strongly agree), and calculate the following seven subscale scores by summing the items on that subscale. Items marked **(R)** are reverse-coded (i.e. 1 = strongly agree and 5 = strongly disagree). To do that, subtract the item response from 6, and use the resulting number as the item score.

1. *Receiving relevant information:* 1, 8, 15, 22, 29, 36, 43, 50, 57
2. *Evaluating the information and comparing it to norms:* 2, 9, 16, 23, 30, 37, 44, 51, 58
3. *Triggering change:* 3, 10, 17, 24, 31, 38, 45, 52, 59
4. *Searching for options:* 4, 11, 18, 25, 32, 39, 46, 53, 60

5. *Formulating a plan:* 5, 12, 19, 26, 33, 40, 47, 54, 61
6. *Implementing the plan:* 6, 13, 20, 27, 34, 41, 48, 55, 62
7. *Assessing the plan's effectiveness:* 7, 14, 21, 28, 35, 42, 49, 56, 63

Based on our clinical and college samples, we tentatively recommend the following ranges for interpreting SRQ total scores with the 63-item scale:

- **>= 239**: High (intact) self-regulation capacity (top quartile)
- **214 - 238**: Intermediate (moderate) self-regulation capacity (middle quartiles)
- **<= 213**: Low (impaired) self-regulation capacity (bottom quartile)

B.5 Metacognition

Metacognitive Awareness Inventory (MAI) proposed by Schraw & Dennison (1994) and revised by Terlecki & McMahon (2018).

*Think of yourself as a **learner**. Read each statement carefully, and rate it as it generally applies to you when you are in the role of a learner (student, attending classes, university etc.) Please indicate how true each reason is for you using the following scale:*

Score	Response
1	I **NEVER** do this
2	I do this **infrequently**
3	I do this **inconsistently**
4	I do this **frequently**
5	I **ALWAYS** do this

1. I ask myself periodically if I am meeting my goals.
2. I consider several alternatives to a problem before I answer.
3. I try to use strategies that have worked in the past.
4. I pace myself while learning in order to have enough time.
5. I understand my intellectual strengths and weaknesses.
6. I think about what I really need to learn before I begin a task.

7. I know how well I did once I finish a test.
8. I set specific goals before I begin a task.
9. I slow down when I encounter important information.
10. I know what kind of information is most important to learn.
11. I ask myself if I have considered all options when solving a problem.
12. I am good at organizing information.
13. I consciously focus my attention on important information.
14. I have a specific purpose for each strategy I use.
15. I learn best when I know something about the topic.
16. I know what the teacher expects me to learn.
17. I am good at remembering information.
18. I use different learning strategies depending on the situation.
19. I ask myself if there was an easier way to do things after I finish a task.
20. I have control over how well I learn.
21. I periodically review to help me understand important relationships.
22. I ask myself questions about the material before I begin.
23. I think of several ways to solve a problem and choose the best one.
24. I summarize what I've learned after I finish.
25. I ask others for help when I don't understand something.
26. I can motivate myself to learn when I need to.
27. I am aware of what strategies I use when I study.
28. I find myself analyzing the usefulness of strategies while I study.
29. I use my intellectual strengths to compensate for my weaknesses.
30. I focus on the meaning and significance of new information.
31. I create my own examples to make information more meaningful.
32. I am a good judge of how well I understand something.
33. I find myself using helpful learning strategies automatically.
34. I find myself pausing regularly to check my comprehension.

35. I know when each strategy I use will be most effective.
36. I ask myself how well I accomplish my goals once I'm finished.
37. I draw pictures or diagrams to help me understand while learning.
38. I ask myself if I have considered all options after I solve a problem.
39. I try to translate new information into my own words.
40. I change strategies when I fail to understand.
41. I use the organizational structure of the text to help me learn.
42. I read instructions carefully before I begin a task.
43. I ask myself if what I'm reading is related to what I already know.
44. I reevaluate my assumptions when I get confused.
45. I organize my time to best accomplish my goals.
46. I learn more when I am interested in the topic.
47. I try to break studying down into smaller steps.
48. I focus on overall meaning rather than specifics.
49. I ask myself questions about how well I am doing while I am learning something new.
50. I ask myself if I learned as much as I could have once I finish a task.
51. I stop and go back over new information that is not clear.
52. I stop and reread when I get confused.

Scoring Directions: Score each response from 1 (never) to 5 (always), and calculate the following subscale scores by summing the items on that subscale.

Knowledge about Cognition:

1. *Declarative Knowledge:* 5, 10, 12, 16, 17, 20, 32, 46 (score out of $8 \times 5 = 40$)
2. *Procedural Knowledge:* 3, 14, 27, 33 (score out of $4 \times 5 = 20$)
3. *Conditional Knowledge:* 15, 18, 26, 29, 35 (score out of $5 \times 5 = 25$)

Regulation of Cognition:

1. *Planning:* 4, 6, 8, 22, 23, 42, 45 (score out of $7 \times 5 = 35$)

2. *Information Management Strategies:* 9, 13, 30, 31, 37, 39, 41, 43, 47, 48 (score out of $10 \times 5 = 50$)
3. *Comprehension Monitoring:* 1, 2, 11, 21, 28, 34, 49 (score out of $7 \times 5 = 35$)
4. *Debugging Strategies:* 25, 40, 44, 51, 52 (score out of $5 \times 5 = 25$)
5. *Evaluation:* 7, 19, 24, 36, 38, 50 (score out of $6 \times 5 = 30$)

C

Questionnaires for Initial `PHASE1` and Final `PHASE3` Phases

Pre-Test session (SES1) is conducted at the beginning of the semester, and the Post-Test session (SES3) is conducted at the end of the semester.

C.1 Pre-Task Questionnaire (for `PHASE1` and `PHASE3`)

(The following items are adapted from Collins-Thompson et al. (2016).)

1. How much do you know about this topic?
   (1) nothing / I know a lot (5)

2. How interested are you to learn more about this topic?
   (1) not at all / very much (5)

3. How difficult do you think it will be to search for information about this topic?
   (1) very easy / very difficult (5)

(The following items are adapted from Crescenzi (2020).)

Indicate your agreement with the following statements.

(1) Strongly Disagree / Neutral / Strongly Agree (5)

4. I am interested to learn more about the topic of this task.
5. I know a lot about this topic.
6. I can write a good summary now without needing to look for information.
7. It will be difficult to determine when I have enough information to write my summary.
8. I think this will be a difficult task.
9. I am confident I know (or can find) adequate information to write a good summary.

C.2 Post-Task Questionnaire (for SES1 and SES3)

(The following items are adapted from Collins-Thompson et al. (2016).)

Indicate your agreement with the following statements.

(1) Not at all / Unlikely / Somewhat / Likely / Very Likely (5)

Search for information exploration:

1. I was cognitively engaged in search task.
2. I made an effort at performing the search task.
3. The time for search was spent productively on meaningful tasks.
4. I was able to explore relationships among multiple concepts.
5. I was able to expand the scope of my knowledge about the topic.
6. I feel that I was able to put together pieces of information into one big concept.

Learner interest and motivation:

7. I feel that I have full understanding of the topic of this task

8. I became more interested in this topic.
9. I would like to find more information about this topic
10. I would like to share what I learned with my people I know.
11. I feel that I learned useful information as a result of this search.
12. I was able to develop new ideas or perspectives.

Perceived learning and search success:

On a scale of 0 - 100

13. How would you grade your learning outcome?
14. How would you grade your search outcome?

(The following items are adapted from Crescenzi (2020).)

Indicate your agreement with the following statements.

(1) Strongly Disagree | Neutral | Strongly Agree (5)

15. Overall, it was difficult to search for information to make the summary.
16. It was difficult to determine search terms to use to find relevant information.
17. It was difficult to decide whether to continue inspecting the search results or to search again.
18. It was difficult to choose which search results to view.
19. It was difficult to determine when to stop looking for information.
20. I would have preferred to think longer about my summary.
21. If I had more time, I would have considered more information.

22. I felt anxious while completing this task.
23. I did not have enough time.
24. It was difficult to decide which sources to select.
25. I felt hurried or rushed during this task.
26. I had adequate information to make a good summary.
27. I felt I had enough information.
28. My understanding of the topic was no longer changing.
29. I collected enough information to make a summary.
30. I was no longer learning about the topic.
31. I felt I had adequate information to make a summary.
32. I was focused on getting information about one thing.
33. I felt continuing the search was a waste of time, as the same information was showing up.
34. I had a list of certain things I was interested in.
35. I stopped searching because I was not finding new information.
36. I stopped searching when I had an option that satisfied the things that were important to me.
37. I only considered looking for the piece of information most important to me.
38. I kept finding the same information in every search.
39. My view of the topic was no longer changing.
40. I was most concerned about finding information on one specific aspect.

C.3 Preference for CTA vs Silent Condition

1. You were asked to talk-aloud for one task, and work in silence for the other. Which one was better?

   - Talk aloud much was a lot better
   - Talk aloud was slightly better
   - I did not feel any difference
   - Working in silence was slightly better
   - Working in silence was a lot better

2. Why? .....

C.4 `PHASE1-SHEG`: Website Reliability Assessment

You are researching children's health and come across this website: https://acpeds.org.

Please decide if this website is a trustworthy source of information on children's health.

You may use any information on this website, or you can open a new tab and do an Internet search if you want. You can consider checking out search engines other than Google, such as Bing, Yahoo, or Ecosia.

Take about 5 minutes to complete this task. Turn YASBIL on before proceeding.

1. Is this website a trustworthy source to learn about children's health? Explain your answer, citing evidence from the webpages you used. Be sure to provide the URLs to the webpages you cite in the next textbox. .....
2. Please paste the URLs to the webpages you used to explain your answer above, one per line. .....

C.5 PHASE3-SHEG: Website Comparison Assessment

Imagine you are doing research on criminal justice reform, and you find the following webpages.

A. https://sites.psu.edu/aspsy/2019/10/20/criminal-justice-reform

B. https://en.wikipedia.org/wiki/Criminal_justice_reform_in_the_United_States

Which of these two webpages do you think is a better place to start your research?

You may use any information on the two webpages, or you can open another tab and do an Internet search if you want. You can consider checking out search engines other than Google, such as Bing, Yahoo, or Ecosia.

Take about 5 minutes to complete this task. Turn YASBIL on before proceeding.

1. Which of these two webpages do you think is a better place to start your research? .....
2. Explain your choice. Refer to both webpages in your answer. .....
3. Please paste the URLs to the webpages you used to explain your answer above, one per line.

D

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

Abualsaud, M., & Smucker, M. D. (2019). Patterns of search result examination: Query to first action. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 1833–1842. https://doi.org/10.1145/3357384.3358041

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.

Arguello, J., & Choi, B. (2019). The effects of working memory, perceptual speed, and inhibition in aggregated search. *ACM Transactions on Information Systems*, *37*(3). https://doi.org/10.1145/3322128

Aula, A., Majaranta, P., & Räihä, K.-J. (2005). Eye-tracking reveals the personal styles for search result evaluation. In M. F. Costabile & F. Paternò (Eds.), *Human-computer interaction - INTERACT 2005* (pp. 1058–1061). Springer Berlin Heidelberg.

Ausubel, D. P. (2012). *The acquisition and retention of knowledge: A cognitive view*. Springer Science & Business Media.

Ausubel, D. P., Novak, J. D., Hanesian, H., et al. (1968). *Educational psychology: A cognitive view* (Vol. 6). Holt, Rinehart; Winston New York.

Bailey, E., & Kelly, D. (2011). Is amount of effort a better predictor of search success than use of specific search tactics? *Proceedings of the American Society for Information Science and Technology*, *48*(1), 1–10.

Balatsoukas, P., & Ruthven, I. (2010). The use of relevance criteria during predictive judgment: An eye tracking approach. *Proceedings of the American Society for Information Science and*

Technology, *47*(1), 1–10. https://doi.org/10.1002/meet.14504701145

Balatsoukas, P., & Ruthven, I. (2012). An eye-tracking approach to the analysis of relevance judgments on the Web: The case of Google search engine. *Journal of the American Society for Information Science and Technology*, *63*(9), 1728–1746. https://doi.org/10.1002/asi.22707

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part i. Background and theory. *Journal of Documentation*.

Beymer, D., Orton, P. Z., & Russell, D. M. (2007). An eye tracking study of how pictures influence online reading. *IFIP Conference on Human-Computer Interaction*, 456–460.

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

Bhattacharya, N., & Gwizdka, J. (2018). Relating eye-tracking measures with changes in knowledge on search tasks. *Symposium on Eye Tracking Research & Applications (ETRA)*.

Bhattacharya, N., & Gwizdka, J. (2019b). Measuring learning during search: Differences in interactions, eye-gaze, and semantic similarity to expert knowledge. *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, 63–71.

Bhattacharya, N., & Gwizdka, J. (2019a). Measuring learning during search: Differences in interactions, eye-gaze, and semantic similarity to expert knowledge. *CHIIR'19*.

Bhattacharya, N., & Gwizdka, J. (2021). YASBIL: Yet another search behaviour (and) interaction logger. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2585–2589.

Bilal, D., & Gwizdka, J. (2016). Children's Eye-fixations on Google Search Results. *Proceedings of the 79th ASIS&T Annual Meeting*, *79*, 89:1–89:6. https://doi.org/10.1002/pra2.2016.14505301089

Blanken-Webb, J. (2017). Metacognition: Cognitive dimensions of e-learning. In *E-learning ecologies* (pp. 163–182). Routledge.

Boldi, P., Bonchi, F., Castillo, C., & Vigna, S. (2009). From" dango" to" japanese cakes": Query reformulation models and patterns. *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, *1*, 183–190.

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

Breakstone, J., McGrew, S., Smith, M., Ortega, T., & Wineburg, S. (2018). Why we need a new approach to teaching digital literacy. *Phi Delta Kappan*, *99*(6), 27–32.

Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., & Saavedra, A. (2021). Students' Civic Online Reasoning: A National Portrait. *Educational Researcher*. https://doi.org/10.3102/0013189X211017495

Broder, A. (2002). A taxonomy of web search. *SIGIR Forum*, *36*(2), 3–10. https://doi.org/10.1145/792550.792552

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

Brown, J. (1998). *Self-regulation and the addictive behaviours*. New York: Plenum Press.

Brown, J. M., Miller, W. R., & Lawendowski, L. A. (1999). The self-regulation questionnaire. In V. L. & J. T. L. (Eds.), *Innovations in clinical practice: A sourcebook* (Vol. 17, pp. 281–292). Professional Resource Press/Professional Resource Exchange.

Buscher, G., Cutrell, E., & Morris, M. R. (2009). What Do You See When You're Surfing? Using Eye Tracking to Predict Salient Regions of Web Pages. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 10.

Buscher, G., Dumais, S. T., & Cutrell, E. (2010). The good, the bad, and the random: An eye-tracking study of ad quality in web search. *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 42–49. https://doi.org/10.1145/1835449.1835459

Chen, Y., Zhao, Y., & Wang, Z. (2020). Understanding online health information consumers' search as a learning process. *Library Hi Tech*.

Cherry, K. (2020). What Is Motivation? In *Verywell Mind*. https://www.verywellmind.com/what-is-motivation-2795378

Cole, L., MacFarlane, A., & Makri, S. (2020). More than words: The impact of memory on how undergraduates with dyslexia interact with information. *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, 353–357. https://doi.org/10.1145/3343413.3378005

Cole, M. J., Gwizdka, J., Liu, C., Belkin, N. J., & Zhang, X. (2013). Inferring user knowledge level from eye movement patterns. *Information Processing & Management*, *49*(5), 1075–1091.

Collins, C. (2021). Reimagining Digital Literacy Education to Save Ourselves. *Learning for Justice*, *Fall 2021*. https://www.learningforjustice.org/magazine/fall-2021/reimagining-digital-literacy-education-to-save-ourselves

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

Collins-Thompson, K., Rieh, S. Y., Haynes, C. C., & Syed, R. (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*, 163–172.

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

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

Crescenzi, A. M. C. (2020). *Adaptation in Information Search and Decision-Making under Time Pressure* [PhD thesis, The University of North Carolina at Chapel Hill University Libraries]. https://doi.org/10.17615/YT6K-AC37

Cutrell, E., & Guan, Z. (2007). What are you looking for? An eye-tracking study of information usage in web search. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 407–416. https://doi.org/10.1145/1240624.1240690

Deci, E. L., & Ryan, R. M. (2013). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.

Dervin, B., & Naumer, C. M. (2010). Sense-making. In M. J. Bates & M. M. N. (Eds.), *Encyclopedia of library and information sciences (3rd ed.)* (pp. 4696--4707). Taylor; Francis.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.

DiCerbo, K. E., & Behrens, J. T. (2014). Impacts of the digital ocean on education. *London: Pearson*, *1*.

Djamasbi, S., Hall-Phillips, A., & Yang, R. (Rachel). (2013). Search Results Pages and Competition for Attention Theory: An Exploratory Eye-Tracking Study. In S. Yamamoto (Ed.), *Human Interface and the Management of Information. Information and Interaction Design* (pp. 576–583). Springer Berlin Heidelberg. http://link.springer.com.ezproxy.lib.utexas.edu/chapter/10.1007/978-3-642-39209-2-64

Dumais, S. T., Buscher, G., & Cutrell, E. (2010). Individual differences in gaze patterns for web search. *Proceedings of the Third Symposium on Information Interaction in Context*, 185–194. https://doi.org/10.1145/1840784.1840812

Egusa, Y., Saito, H., Takaku, M., Terai, H., Miwa, M., & Kando, N. (2010). Using a Concept Map to Evaluate Exploratory Search. *Proceedings of the Third Symposium on Information Interaction in Context*, 175–184. https://doi.org/10.1145/1840784.1840810

Egusa, Y., Takaku, M., & Saito, H. (2014a). How Concept Maps Change if a User Does Search or Not? *Proceedings of the 5th Information Interaction in Context Symposium*, 68–75. https://doi.org/10.1145/2637002.2637012

Egusa, Y., Takaku, M., & Saito, H. (2014b). How to evaluate searching as learning. *Searching as Learning Workshop (IIiX 2014 Workshop)*. http://www.diigubc.ca/IIIXSAL/program.html

Egusa, Y., Takaku, M., & Saito, H. (2017). Evaluating Complex Interactive Searches Using Concept Maps. *SCST@ CHIIR*, 15–17.

Eickhoff, C., Dungs, S., & Tran, V. (2015). An eye-tracking study of query reformulation. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 13–22. https://doi.org/10.1145/2766462.2767703

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.

Eickhoff, C., Teevan, J., White, R., & Dumais, S. (2014). Lessons from the journey: A query log analysis of within-session learning. *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, 223–232.

Fleischmann, K., Verma, N., Gursoy, A., Bautista, J. R., & Day, J. (2022). I 303 : Ethical foundations for informatics [syllabus]. *School of Information, University of Texas at Austin*.

Francis, G., MacKewn, A., & Goldthwaite, D. (2004). *CogLab on a CD*. Wadsworth Publishing Company.

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.

Gadiraju, U., Yu, R., Dietze, S., & Holtz, P. (2018). Analyzing knowledge gain of users in informational search sessions on the web. *Conference on Human Information Interaction & Retrieval (CHIIR)*.

Ghosh, S., Rath, M., & Shah, C. (2018). Searching as learning: Exploring search behavior and learning outcomes in learning-related tasks. *Conference on Human Information Interaction & Retrieval (CHIIR)*.

Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., & Wichansky, A. M. (2002). Eye tracking in web search tasks: Design implications. *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, 51–58.

González-Ibáñez, R., Esparza-Villamán, A., Vargas-Godoy, J. C., & Shah, C. (2019). A comparison of unimodal and multimodal models for implicit detection of relevance in interactive IR. *Journal of the Association for Information Science and Technology*, *0*(0). https://doi.org/10.1002/asi.24202

Gossen, T., Höbel, J., & Nürnberger, A. (2014). A comparative study about children's and adults' perception of targeted web search engines. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1821–1824. https://doi.org/10.1145/2556288.2557031

Grabowski, B. L. (1996). Generative learning: Past, present, and future. *Handbook of Research for Educational Communications and Technology*, 897–918.

Granka, L. A., Joachims, T., & Gay, G. (2004). Eye-tracking analysis of user behavior in WWW search. *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 478–479. https://doi.org/10.1145/1008992.1009079

Groner, R., Walder, F., & Groner, M. (1984). Looking at faces: Local and global aspects of scanpaths. In *Advances in psychology* (Vol. 22, pp. 523–533). Elsevier.

Guan, Z., & Cutrell, E. (2007). An eye tracking study of the effect of target rank on web search. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 417–420. https://doi.org/10.1145/1240624.1240691

Guyan, M. (2013). Improving Learner Motivation for eLearning. In *Learning Snippets*. https://learningsnippets.wordpress.com/2013/10/30/improving-learner-motivation-for-elearning/

Gwizdka, J. (2013). Effects of working memory capacity on users' search effort. *Proceedings of the International Conference on Multimedia, Interaction, Design and Innovation*, 11:1–11:8. https://doi.org/10.1145/2500342.2500358

Gwizdka, J. (2014). Characterizing Relevance with Eye-tracking Measures. *Proceedings of the 5th Information Interaction in Context Symposium*, 58–67. https://doi.org/10.1145/2637002.2637011

Gwizdka, J. (2017). I Can and So I Search More: Effects Of Memory Span On Search Behavior. *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, 341–344. https://doi.org/10.1145/3020165.3022148

Gwizdka, J. (2018). Inferring Web Page Relevance Using Pupillometry and Single Channel EEG. In F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, & A. B. Randolph (Eds.), *Information Systems and Neuroscience* (pp. 175–183). Springer International Publishing. https://doi.org/10.1007/978-3-319-67431-5-20

Gwizdka, J., & Bilal, D. (2017). Analysis of Children's Queries and Click Behavior on Ranked Results and Their Thought Processes in Google Search. *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, 377–380. https://doi.org/10.1145/3020165.3022157

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.

Gwizdka, J., & Zhang, Y. (2015a). Differences in Eye-Tracking Measures Between Visits and Revisits to Relevant and Irrelevant Web Pages. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 811–814. https://doi.org/10.1145/2766462.2767795

Gwizdka, J., & Zhang, Y. (2015b). Towards Inferring Web Page Relevance – An Eye-Tracking Study. *Proceedings of iConference'2015*, 5. https://www.ideals.illinois.edu/handle/2142/73709

Halttunen, K., & Jarvelin, K. (2005). Assessing learning outcomes in two information retrieval learning environments. *Information Processing & Management*, *41*(4), 949–972. https://doi.org/10.1016/j.ipm.2004.02.004

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

Hassan, A., White, R. W., Dumais, S. T., & Wang, Y.-M. (2014). Struggling or exploring? Disambiguating long search sessions. *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, 53–62.

He, J., Qvarfordt, P., Halvey, M., & Golovchinsky, G. (2016). Beyond actions: Exploring the discovery of tactics from user logs. *Information Processing & Management*, *52*(6), 1200–1226.

Hofmann, K., Mitra, B., Radlinski, F., & Shokouhi, M. (2014). An eye-tracking study of user interactions with query auto completion. *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, 549–558. https://doi.org/10.1145/2661829.2661922

Huang, X., & Soergel, D. (2013). Relevance: An improved framework for explicating the notion. *Journal of the American Society for Information Science and Technology*, *64*(1), 18–35. https://doi.org/10.1002/asi.22811

Jiang, J., He, D., & Allan, J. (2014). Searching, browsing, and clicking in a search session: Changes in user behavior by task and over time. *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, 607–616.

https://doi.org/10.1145/2600428.2609633

Josephson, S., & Holmes, M. E. (2002). Visual attention to repeated internet images: Testing the scanpath theory on the world wide web. *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, 43–49.

Jossberger, H., Brand-Gruwel, S., Boshuizen, H., & Van de Wiel, M. (2010). The challenge of self-directed and self-regulated learning in vocational education: A theoretical analysis and synthesis of requirements. *Journal of Vocational Education and Training*, *62*(4), 415–440.

Kahne, J., Lee, N.-J., & Feezell, J. T. (2012). Digital media literacy education and online civic and political participation. *International Journal of Communication*, *6*, 24.

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

Kanfer, F. H. (1970a). *Self-monitoring: Methodological limitations and clinical applications.*

Kanfer, F. H. (1970b). Self-regulation: Research, issues, and speculations. *Behavior Modification in Clinical Psychology*, *74*, 178–220.

Kanniainen, L., Kiili, C., Tolvanen, A., Aro, M., Anmarkrud, Ø., & Leppänen, P. H. T. (2021). Assessing reading and online research comprehension: Do difficulties in attention and executive function matter? *Learning and Individual Differences*, *87*, 101985. https://doi.org/10.1016/j.lindif.2021.101985

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. (2009). Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*, *3*(1—2), 1–224.

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

Knowles, M. S. (1975). *Self-directed learning: A guide for learners and teachers.* New York: Association press.

Ko, A. J. (2021). Seeking information. In *Foundations of Information*. https://faculty.wa shington.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/.

Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., González Perilli, F., Pires, A., Vilaro, A., & Villalobos, N. (2015). Gaze transition entropy. *ACM Transactions on Applied Perception (TAP)*, *13*(1), 1–20.

Krejtz, K., Szmidt, T., Duchowski, A. T., & Krejtz, I. (2014). Entropy-based statistical analysis of eye movement transitions. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 159–166.

Kruikemeier, S., Lecheler, S., & Boyer, M. M. (2018). Learning from news on different media platforms: An eye-tracking experiment. *Political Communication*, *35*(1), 75–96.

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

Lam, H., Russell, D., Tang, D., & Munzner, T. (2007). Session viewer: Visual exploratory analysis of web session logs. *2007 IEEE Symposium on Visual Analytics Science and Technology*, 147–154.

Leacock, C., & Chodorow, M. (1998). Combining local context and WordNet similarity for word sense identification. *WordNet: An Electronic Lexical Database*, *49*(2), 265–283.

Lei, P.-L., Sun, C.-T., Lin, S. S., & Huang, T.-K. (2015). Effect of metacognitive strategies and verbal-imagery cognitive style on biology-based video search and learning performance. *Computers & Education*, *87*, 326–339.

Leu, D. J., Forzani, E., Rhoads, C., Maykel, C., Kennedy, C., & Timbrell, N. (2015). The New Literacies of Online Research and Comprehension: Rethinking the Reading Achievement Gap. *Reading Research Quarterly*, *50*(1), 37–59. https://doi.org/10.1002/rrq.85

Li, Y., & Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management*, *44*(6), 1822–1837.

Ling, C., Steichen, B., & Choulos, A. G. (2018). A comparative user study of interactive multilingual search interfaces. *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, 211–220. https://doi.org/10.1145/3176349.3176383

Liu, C., Gwizdka, J., Liu, J., Xu, T., & Belkin, N. J. (2010). Analysis and evaluation of query reformulations in different task types. *Proceedings of the American Society for Information Science and Technology*, *47*(1), 1–9.

Liu, Z., Liu, Y., Zhou, K., Zhang, M., & Ma, S. (2015). Influence of vertical result in web search examination. *Proceedings of the 38th International ACM SIGIR Conference on*

Research and Development in Information Retrieval, 193–202. https://doi.org/10.1145/2766462.2767714

Lorigo, L., Haridasan, M., Brynjarsdóttir, H., Xia, L., Joachims, T., Gay, G., Granka, L., Pellacini, F., & Pan, B. (2008). Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology*, *59*(7), 1041–1052. https://doi.org/10.1002/asi.20794

Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L., & Gay, G. (2006). The influence of task and gender on search and evaluation behavior using google. *Information Processing & Management*, *42*(4), 1123–1131.

Loyens, S. M. M., Magda, J., & Rikers, R. M. J. P. (2008). Self-Directed Learning in Problem-Based Learning and its Relationships with Self-Regulated Learning. *Educational Psychology Review*, *20*(4), 411–427. https://doi.org/10.1007/s10648-008-9082-7

Mannion, J. (2020). Metacognition, self-regulation and self-regulated learning: What's the difference? In *impact.chartered.college*. https://impact.chartered.college/article/metacognition-self-regulation-regulated-learning-difference/

Mao, J., Liu, Y., Kando, N., Zhang, M., & Ma, S. (2018). How does domain expertise affect users' search interaction and outcome in exploratory search? *ACM Transactions on Information Systems*, *36*.

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.

Marton, F., & Säaljö, R. (1976). On qualitative differences in learning—ii outcome as a function of the learner's conception of the task. *British Journal of Educational Psychology*, *46*(2), 115–127.

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—outcome and process. *British Journal of Educational Psychology*, *46*(1), 4–11.

McGrew, S. (2020). Learning to evaluate: An intervention in civic online reasoning. *Computers & Education*, *145*, 103711.

McGrew, S. (2021). Skipping the source and checking the contents: An in-depth look at students' approaches to web evaluation. *Computers in the Schools*, *38*(2), 75–97.

McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education*, *46*(2), 165–193.

McGrew, S., & Glass, A. C. (2021). Click Restraint: Teaching Students to Analyze Search Results. *Proceedings of the 14th International Conference on Computer-Supported Collaborative Learning-CSCL 2021*.

McGrew, S., Ortega, T., Breakstone, J., & Wineburg, S. (2017). The challenge that's bigger than fake news: Civic reasoning in a social media environment. *American Educator*, *41*(3), 4.

Mihailidis, P., & Thevenin, B. (2013). Media literacy as a core competency for engaged citizenship in participatory democracy. *American Behavioral Scientist*, *57*(11), 1611–1622.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

Miller, W. R., & Brown, J. M. (1991). Self-regulation as a conceptual basis for the prevention and treatment of addictive behaviours. *Self-Control and the Addictive Behaviours*, 3–79.

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.

Next Generation Science Standards. (2021). *Task annotation project in science / sense-making*. https://www.nextgenscience.org/sites/default/files/TAPS%20Sense-making.pdf.

Novak, J. D. (2002). Meaningful learning: The essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science Education*, *86*(4), 548–571.

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

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge University Press. https://doi.org/10.1017/CBO9781139173469

O'Brien, H. L., Kampen, A., Cole, A. W., & Brennan, K. (2020). The role of domain knowledge in search as learning. *Conference on Human Information Interaction and Retrieval (CHIIR)*.

Palani, S., Fourney, A., Williams, S., Larson, K., Spiridonova, I., & Morris, M. R. (2020). An eye tracking study of web search by people with and without dyslexia. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 729–738. https://doi.org/10.1145/3397271.3401103

Pan, B., Hembrooke, H. A., Gay, G. K., Granka, L. A., Feusner, M. K., & Newman, J. K. (2004). The determinants of web page viewing behavior: An eye-tracking study. *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*, 147–154.

Pea, R., & Jacks, D. (2014). *The learning analytics workgroup: A report on building the field of learning analytics for personalized learning at scale.* https://ed.stanford.edu/sites/default/files/law-report-complete-09-02-2014.pdf; Stanford, CA: Stanford University.

Pennanen, M., & Vakkari, P. (2003). Students' conceptual structure, search process, and outcome while preparing a research proposal: A longitudinal case study. *Journal of the American Society for Information Science and Technology*, *54*(8), 759–770.

Piaget, J. (1936). *Origins of intelligence in children.*

Pirolli, P., Schank, P., Hearst, M., & Diehl, C. (1996). Scatter/gather browsing communicates the topic structure of a very large text collection. *Conference on Human Factors in Computing Systems (CHI'96).*

Qvarfordt, P., Golovchinsky, G., Dunnigan, T., & Agapie, E. (2013). Looking ahead: Query preview in exploratory search. *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 243–252. https://doi.org/10.1145/2484028.2484084

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

Rieh, S. Y., Kim, Y.-M., & Markey, K. (2012). Amount of invested mental effort (AIME) in online searching. *Information Processing & Management*, *48*(6), 1136–1150.

Roy, N., Moraes, F., & Hauff, C. (2020). Exploring users' learning gains within search sessions. *Conference on Human Information Interaction and Retrieval (CHIIR).*

Roy, N., Torre, M. V., Gadiraju, U., Maxwell, D., & Hauff, C. (2021). Note the highlight: Incorporating active reading tools in a search as learning environment. *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, 229–238.

Rumelhart, D. E., & Norman, D. A. (1981). Accretion, tuning and restructuring: Three modes of learning. In J. W. Cotton & K. R. (Eds.), *Semantic factors in cognition* (pp. 37–90).

Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, S. R. J., & M. W. E. (Eds.), *Schooling and the acquisition of knowledge* (pp. 99–135). Hillsdale, NJ: Erlbaum.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, *43*(3), 450.

Ryan, R. M., & Deci, E. L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67.

Ryan, R. M., & Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.

Saks, K., & Leijen, Ä. (2014). Distinguishing Self-directed and Self-regulated Learning and Measuring them in the E-learning Context. *Procedia - Social and Behavioral Sciences*, *112*, 190–198. https://doi.org/10.1016/j.sbspro.2014.01.1155

Saracevic, T. (1975). Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, *26*(6), 321–343.

Saracevic, T. (2007a). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology*, *58*(13), 1915–1933. https://doi.org/10.1002/asi.20682

Saracevic, T. (2007b). 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*(13), 2126–2144.

Saracevic, T. (2016). The Notion of Relevance in Information Science: Everybody knows what relevance is. But, what is it really? *Synthesis Lectures on Information Concepts, Retrieval, and Services*.

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

Scharinger, C., Kammerer, Y., & Gerjets, P. (2016). Fixation-Related EEG Frequency Band Power Analysis: A Promising Neuro-Cognitive Methodology to Evaluate the Matching-Quality of Web Search Results? *HCI International 2016 – Posters' Extended Abstracts*, 245–250. https://doi.org/10.1007/978-3-319-40548-3-41

Schraw, G., & Dennison, R. S. (1994). Assessing Metacognitive Awareness. *Contemporary Educational Psychology*, *19*(4), 460–475. https://doi.org/10.1006/ceps.1994.1033

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129.

Slanzi, G., Balazs, J. A., & Velásquez, J. D. (2017). Combining eye tracking, pupil dilation and EEG analysis for predicting web users click intention. *Information Fusion*, *35*, 51–57.

https://doi.org/10.1016/j.inffus.2016.09.003

Smith, C. L., Gwizdka, J., & Feild, H. (2016). Exploring the use of query auto completion: Search behavior and query entry profiles. *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, 101–110. https://doi.org/10.1145/2854946.2854975

Smith, C. L., & Kantor, P. B. (2008). User adaptation: Good results from poor systems. *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 147–154.

Spink, A. (1997). Study of interactive feedback during mediated information retrieval. *Journal of the American Society for Information Science*.

Stanford History Education Group (SHEG). (2021a). *Webpage comparison | Civic Online Reasoning*. https://cor.stanford.edu/curriculum/assessments/webpage-comparison.

Stanford History Education Group (SHEG). (2021b). *Website Reliability | Civic Online Reasoning*. https://cor.stanford.edu/curriculum/assessments/website-reliability.

Syed, R., Collins-Thompson, K., Bennett, P. N., Teng, M., Williams, S., Tay, D. W. W., & Iqbal, S. (2020). Improving learning outcomes with gaze tracking and automatic question generation. *The Web Conference (WWW)*.

Tahamtan, I. (2019). The effect of motivation on web search behaviors of health consumers. *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, 401–404.

Taramigkou, M., Apostolou, D., & Mentzas, G. (2018). Leveraging exploratory search with personality traits and interactional context. *Information Processing & Management*, *54*(4), 609–629.

Terlecki, M. (2020). Revising the Metacognitive Awareness Inventory (MAI) to be More User-Friendly. In *Improve with Metacognition*. https://www.improvewithmetacognition.com/revising-the-metacognitive-awareness-inventory/

Terlecki, M., & McMahon, A. (2018). A Call for Metacognitive Intervention: Improvements Due to Curricular Programming in Leadership. *Journal of Leadership Education*, *17*(4), 130–145. https://doi.org/10.12806/V17/I4/R8

Urgo, K., & Arguello, J. (2022). Learning assessments in search-as-learning: A survey of prior work and opportunities for future research. *Information Processing & Management*, *59*(2), 102821.

Vakkari, P. (2000). Cognition and changes of search terms and tactics during task performance: A longitudinal case study. In *Content-based multimedia information access-volume 1* (pp. 894–907).

Vakkari, P. (2001a). 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.

Vakkari, P. (2001b). A theory of the task-based information retrieval process: A summary and generalisation of a longitudinal study. *Journal of Documentation*, *57*(1), 44–60. https://doi.org/10.1108/EUM0000000007075

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

Vallat, R. (2018). Pingouin: Statistics in python. *J. Open Source Softw.*, *3*(31), 1026.

Vancouver Island University. (2021). *Metacognitive awareness inventory (MAI)*. https://alearningjourneyweb.files.wordpress.com/2017/03/self-directed-vs-self-regulated-chart.pdf

Villa, R., & Halvey, M. (2013). Is relevance hard work? Evaluating the effort of making relevant assessments. *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 765–768.

Wang, Y., Yin, D., Jie, L., Wang, P., Yamada, M., Chang, Y., & Mei, Q. (2018). Optimizing whole-page presentation for web search. *ACM Trans. Web*, *12*(3). https://doi.org/10.1145/3204461

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

Wilson, M. J., & 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), 291–306.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, *55*(3), 249–270.

Wineburg, S., & McGrew, S. (2016). Why students can't google their way to the truth. *Education Week*, *36*(11), 22–28.

Wineburg, S., & McGrew, S. (2017). *Lateral reading: Reading less and learning more when evaluating digital information*.

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, *24*(4), 345–376.

Xu, L., Zhou, X., & Gadiraju, U. (2020). How does team composition affect knowledge gain of users in collaborative web search? *Conference on Hypertext and Social Media (HT)*.

Yu, R., Gadiraju, U., Holtz, P., Rokicki, M., Kemkes, P., & Dietze, S. (2018). Predicting User Knowledge Gain in Informational Search Sessions. *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 75–84. https://doi.org/10.1145/3209978.3210064

Zhang, P., & Soergel, D. (2014). Towards a comprehensive model of the cognitive process and mechanisms of individual sensemaking. *Journal of the Association for Information Science and Technology*, *65*(9), 1733–1756. https://doi.org/10.1002/asi.23125

Zhang, P., & Soergel, D. (2016). Process patterns and conceptual changes in knowledge representations during information seeking and sensemaking: A qualitative user study. *Journal of Information Science*, *42*(1), 59–78.

Zhang, X., Cole, M., & Belkin, N. (2011). Predicting Users' Domain Knowledge from Search Behaviors. *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1225–1226. https://doi.org/10.1145/2009916.2010131

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, *81*(3), 329.

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