

A Characterization of Early Career Researchers' Activities With Academic Sources  
for Writing

by

Mohammad Shakirul Islam  
Bachelor of Science in Computer Science and Engineering  
Daffodil International University, 2019

A Thesis Submitted in Partial Fulfillment of the  
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science

© Mohammad Shakirul Islam, 2024  
University of Victoria

All rights reserved. This dissertation may not be reproduced in whole or in part, by  
photocopying or other means, without the permission of the author.

A Characterization of Early Career Researchers' Activities With Academic Sources  
for Writing

by

Mohammad Shakirul Islam  
Bachelor of Science in Computer Science and Engineering  
Daffodil International University, 2019

Supervisory Committee

---

Dr. Miguel Nacenta, Supervisor  
(Department of Computer Science)

---

Dr. Neil Ernst, Departmental Member  
(Department of Computer Science)

## ABSTRACT

Research processes involve gathering and understanding previous work, including finding existing literature, reading, annotating, collecting and, finally, synthesizing summaries as literature surveys and related work sections. Although much work has been devoted to understanding how people find and retrieve previous work and how people use existing reference managers, in this paper we consider the processes involving research of previous literature in a broader context, looking at how the found sources and references are incorporated into the process of writing new papers and reports.

We propose an activity model (RaMSeS) for the general process of research with academic sources; from paper search to the writing of text that incorporates previous work citations. We used this model to design a survey that investigates current practices by early career researchers. Through the survey, we were able to classify early career researchers into three coarse groups: casual collection managers (who use reference management systems less and tend not to revisit their collections), traditional document managers (who tend to take comprehensive notes but use multiple systems to manage information), and digitally-savvy collection managers (who are more interested in organizing and categorizing their document collections). We also learned about the ways in which participants use their source collections for writing, refreshing their knowledge, and recognizing patterns in the literature. We also conducted an evaluation study with experts in the field of library sciences to understand the applicability of our proposed model (RaMSeS) to assist in the teaching and explanation of the entire process of collecting and curating bibliographies for early career researchers. We provide a thorough thematic analysis of the interviews that we conducted with the experts who overall, found the RaMSeS model useful and provided insightful feedback on how it can be applied to the entire research process. Our findings can support the development of tools that further support the later parts of the research process when existing literature is re-read and analyzed to become part of new research documents.

# Contents

<b>Supervisory Committee</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>Contents</b>	<b>iv</b>
<b>List of Tables</b>	<b>vii</b>
<b>List of Figures</b>	<b>viii</b>
<b>Acknowledgements</b>	<b>ix</b>
<b>Dedication</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	4
1.2 Scope . . . . .	5
1.3 Organization of the thesis . . . . .	6
<b>2 Background and Related Work</b>	<b>8</b>
2.1 Information Seeking and Retrieval . . . . .	8
2.2 Sensemaking and Searching Practices of Knowledge Practitioners . .	9
2.3 Reference Management Systems (RMS) . . . . .	10
2.4 Information Literacy (IL) Frameworks . . . . .	12
<b>3 Researcher’s Management of Search and Storage (RaMSeS) Model</b>	<b>14</b>
3.1 Description of RaMSeS Terms . . . . .	17
3.2 Alice’s Journey through the RaMSeS Model . . . . .	19
3.3 Comparison of RaMSeS to other models . . . . .	20
<b>4 Validation of the RaMSeS Model</b>	<b>27</b>

4.1	Interview Research Questions . . . . .	27
4.2	Interview Design . . . . .	28
4.3	Participant Recruitment . . . . .	28
4.4	Study Procedure . . . . .	29
4.5	Data Collection and Analysis . . . . .	31
4.6	Interview Findings . . . . .	31
4.6.1	Literature Review Phases . . . . .	33
4.6.2	The Dynamic Nature of the Research Process . . . . .	35
4.6.3	Areas of Focus in the Literature Review Process . . . . .	35
4.6.4	Effectiveness of Handling the Research Processes . . . . .	36
4.6.5	Challenges in Literature Review Process . . . . .	38
4.6.6	Models and Frameworks mentioned by Participants . . . . .	39
4.6.7	Representation and Applicability of the RaMSeS Model . . . . .	40
4.6.8	The Usefulness of the RaMSeS Model on Teaching Research Methodology . . . . .	41
4.6.9	Suggested Improvements for Enhancing the RaMSeS Model as a Teaching Tool . . . . .	42
4.6.10	Techniques and Tools for Organizing and Managing Literature Findings: Insights from Participants . . . . .	44
4.7	Summary and Discussion . . . . .	45
<b>5</b>	<b>Early Career Researchers' Management of Academic Sources</b>	<b>48</b>
5.1	Survey Research Questions . . . . .	48
5.2	Survey Design . . . . .	49
5.3	Linking Survey Questions to RaMSeS and Research Questions . . . . .	49
5.4	Participant Recruitment . . . . .	51
5.5	Survey Results . . . . .	52
5.5.1	Overall Survey Answers . . . . .	53
5.5.2	Correlation Analysis . . . . .	56
5.5.3	Unsupervised Clustering Analysis . . . . .	58
5.5.4	Open-ended Survey Response Analysis . . . . .	61
5.6	Summary and Discussion . . . . .	62
<b>6</b>	<b>Discussion and Future Directions</b>	<b>66</b>
6.1	Future Directions . . . . .	68

<b>7 Conclusion</b>	<b>69</b>
<b>Bibliography</b>	<b>71</b>
<b>A Additional Information</b>	<b>78</b>

# List of Tables

Table 3.1 Comparison of RaMSeS with Other Models . . . . .	23
Table 4.1 Interview Plan and Questions . . . . .	30
Table 4.2 Interview Codebook . . . . .	32
Table 5.1 Survey questions and associated question types. . . . .	51
Table 5.2 Overall demographics of our Participants . . . . .	52

# List of Figures

Figure 3.1	Researcher’s Management of Search and Storage (RaMSeS) Model	15
Figure 3.2	Bloom’s Taxonomy [4]	20
Figure 3.3	The Big6 Information Literacy Model [26]	21
Figure 3.4	Ellis’s Model of information seeking Behavior [69]	22
Figure 3.5	Kuhlthau’s Information Search Process [57]	23
Figure 3.6	Wilson’s Model of Information Behaviour [69]	24
Figure 3.7	The ACRL Framework for Information Literacy [66]	25
Figure 5.1	Survey Results: The answers to ranking questions (SQ9, SQ12, and SQ20) are in ascending order, with the left choices having lower cumulative averages.	54
Figure 5.2	Correlation matrix visualization with Pearson correlation coefficients, annotated with statistically significant values.	57
Figure 5.3	Visualization of a 3-cluster (k=3) analysis highlighting the top 10 significant features, ranked by importance. Survey responses were encoded as binary or numeric values. The y-axis shows each cluster’s average tendency to choose an option, with the x-axis set at a baseline of 0, indicating the overall average.	59

## ACKNOWLEDGEMENTS

I would like to thank:

**Dr. Miguel Nacenta** for his constant support in my academic and research areas, providing me with valuable feedback on my research, which helped me improve my potential to accomplish my goals. I would also like to thank him for allowing me to pursue co-op work with the Ministry of Health at the Government of British Columbia. It provided me with great opportunities and opened doors for my future career. I am very grateful for his belief in me, which enabled me to finish my research successfully and move forward.

**Hamid Mansoor** for his endless support in my research journey, providing me with his time and knowledge. Throughout my research design, analysis, and writing, he was a mentor who guided me with inspiration and feedback.

**My Wife** for being my constant source of motivation and encouragement during this journey. She stood by me at each stage of hardship and also provided financial support during my student tenure. Her trust in me was the inspiration to finish this degree along with two consecutive co-op terms.

**My Parents** for always being supportive whenever needed. Without their tremendous support, love, and encouragement, I would not be where I am today.

Finally, I would like to thank all of my VIXI research group members for their assistance and inspiration throughout my master's program.

*Read in the name of your Lord who created, created man from clot of blood. Read, your Lord is the most Bounteous, Who has taught the use of the pen, has taught man what he did not know.*

Quran, 96:1-5

## DEDICATION

To my beloved wife, Arafin Sultana, my mother, Samsunnahar Romisa and my father, Mohammad Ruhul Amin.

# Chapter 1

## Introduction

Academic researchers are often engaged in the task of managing large-scale document collections. These collections, encompassing an array of scholarly articles and research papers, serve as a foundation for conducting literature reviews, generating new ideas, and crafting high-quality academic reports and papers. Given the importance of such collections, extant work has reported on the process of organizing large-scale academic information collection [52, 71, 60, 15]. However, such work has mostly focused on Reference Management Systems (RMS) and researchers' habits and preferences when using such systems [65]. Although important, this offers only a limited view of a process of research that includes other additional activities that go beyond just collecting references. For example, researchers often engage in reviewing previously read work (revisiting), organizing and classifying existing work, and using the reference database itself to compose the final documents (e.g., a related work section). Proske et. al. [54] highlight that most work focuses on just the process of building collections and not enough on *how* these collections are used, and especially re-used. In addition, current work often includes only participants from a single institution [49, 58].

Existing work that deals with understanding researchers' reference management has mostly focused on specific tool usage and specific methodology usage. However, our focus is on describing the reference management process, encompassing the entire research journey from the early stages of background knowledge acquisition to the final report writing. To describe this comprehensive research process, we devised the Researcher's Management of Search and Storage (RaMSeS) model, which attempts to simulate the entire academic research process in traditional academic settings. Unlike previous works that concentrated on individual aspects of the research process, RaMSeS offers a detailed framework for managing search and storage activities across

all stages. This model not only provides a structured approach to understanding the research process but also holds potential as an educational aid for guiding early-career researchers. Using RaMSeS as our foundation, we designed and conducted two studies to validate and explore current research practices.

In this thesis, we first described our Researcher's Management of Search and Storage (RaMSeS) model. Using the model, we then designed a survey study to capture the current practices of early career researchers. The survey study was not focused on evaluating RMS but on understanding and describing the entire research journey. The results of the survey encouraged us to validate the proposed model, leading to the design of the interview study. For coherence, the thesis is organized to first introduce the model, followed by its validation, and finally, the survey findings. It is important to note that our study is specifically focused on traditional academic research contexts and researchers who use these systems for organizing and referencing academic papers. Consequently, our findings may not fully address research methodologies outside of traditional literature-based approaches, such as community-based research or indigenous ways of learning.

In the interview study, we recruited experts in managing research literature processes to validate the RaMSeS model of information seeking behaviour. We aimed to gain a clearer understanding of the information seeking process, particularly literature review building and seeking out research materials. This study involved semi-structured interviews to provide a more granular understanding of the research process as conducted by early career researchers. Through this qualitative approach, we refined our initial characterizations to enhance the practical relevance of our findings, enabling a more effective support framework for managing academic literature.

The research questions of the *interview study*, inquire about the effectiveness of our proposed RaMSeS model. The research questions were built on the insights gathered from our survey study and focused on validating the RaMSeS model and understanding its effectiveness in teaching research methodology. Specifically, the questions addressed whether the RaMSeS model comprehensively represents the key activities and sequences undertaken by literature reviewers in academic research, to what extent experts in research literature processes find the RaMSeS model useful, and what are the primary limitations or missing aspects of the model. Interview data analysis shows that participants viewed the RaMSeS model as a comprehensive and flexible representation for describing literature review process, effectively mirroring academic literature management methodologies. The RaMSeS model was also seen as

a valuable tool for teaching research methodology, particularly for librarians who assist students with research. However, participants noted several limitations, including the need for a clearer presentation of the model to enhance its understandability.

In our survey study, we aimed to enhance current insights into the utilization of RMS by expanding the scope of analysis to encompass the broader domain of composing documents reliant on references. Additionally, we sought to increase geographical diversity in our sample by deploying an online survey to crowd-source participants from around the world. Our approach was guided by the RaMSeS model, which outlines the entire process of managing academic document collections as part of creating academic documents. The RaMSeS model facilitates the identification and specification of distinct activities constituting the entire process, as our focus extends beyond just the use of RMS. Furthermore, it furnishes a valuable lexicon for formulating the questions for our survey study, thereby enabling an exploration into the research practices of early career researchers.

In the survey, we targeted graduate students, postdoctoral researchers, and faculty members to capture their research material collection and curation habits and the challenges they face in their respective research processes. The research questions of the survey study inquired about habits such as whether researchers read individual articles before collecting them, how and when they take notes or summarize, and how they revise their collections. We also examine the extent to which existing tools (e.g., RMS like Zotero, PDF files, pen and paper, etc.) support researchers in these tasks. The findings derived from the hundred responses to the online questionnaire can help in understanding early career academic researchers reading and note-taking habits and research collection management approaches. Given the complexity of the data gathered, we utilized unsupervised clustering to create coarse groupings of the participants that provided insights into their research process. We provided a detailed description of these groupings in the survey study chapter 5.

The main contributions of this work are:

- Researcher's Management of Search and Storage (RaMSeS): A model to illustrate the researchers' journey from information seeking/synthesis to final report/result writing.
- Validation of the RaMSeS Model: We recruited five experts in managing research literature processes who recognized RaMSeS as a comprehensive framework for describing literature review process. They also mentioned that the

RaMSeS model may serve as a valuable tool for teaching academic research methodology.

- Results from an online survey with one hundred early career academic researchers to capture their research material collection and curation habits and the challenges they face in their respective research processes.
- Discussion and identification of research directions and challenges for academic document collection creation and curation.

## 1.1 Motivation

The motivation behind this investigation stems from the realization of how important effective handling of scholarly sources is to the process of scholarly writing and research. Effective source management is crucial because it builds on existing knowledge, ensures accurate citation, and prevents plagiarism upholds the integrity and credibility of academic work [12]. Even with the widespread availability of advanced RMS and digital tools, researchers particularly those just starting in their careers often struggle to efficiently select, retrieve, and utilize their informational repositories [54].

The goal of this research is to lessen the gap that exists between the promise of current technological solutions and the actual needs and practices of researchers. This research attempts to contribute to the development of more user-friendly and helpful tools that align with users' workflows and enhance research efficacy by exploring how early career researchers engage with their academic materials.

We broaden the investigation of RMS by delving into the extensive range of activities associated with academic source management among early career researchers. Going beyond the traditional focus on RMS utility and adoption the study explores additional factors influencing user engagement [72]. We aimed to uncover the intricate ways researchers interact with their academic sources from acquisition to annotation and utilization in their writing, and how digital tools are integrated into these processes. We covered the various stages of source management, seeking to provide an in-depth understanding of the practices and challenges faced by researchers. Also, we tried to identify opportunities for technological and methodological improvements to enhance their scholarly work.

Furthermore, the goal of this research is to add to the body of knowledge already available to scholars about information seeking behaviour by offering a more complex story that goes beyond the use of RMSs. The study intends to inform future instruments and approaches that promote more efficient and productive research practices by identifying the many strategies and routines researchers use to manage their academic literature.

## 1.2 Scope

This section outlines the scope of the research, focusing on model development, target participants, and target audience.

**Model Development:** The creation of the Researcher’s Management of Search and Storage (RaMSeS) Model, a framework to explain the processes researchers employ in managing and utilizing academic literature. The model is meant to describe the process, not prescribe any specific path for the research process. The RaMSeS model, derived from an extensive review of existing literature and interdisciplinary insights, outlines the comprehensive journey of academic literature management from resource identification to application in scholarly work. Aimed at aligning with researchers’ real-world practices, the model may provide a systematic structure for examining and enhancing academic literature management processes. Developed through continuous refinement, RaMSeS aspires to advance both research and practices in academic information management. While we recognize George Box’s famous aphorism “*all models are wrong but some are useful*” [13], we are hopeful that the RaMSeS model might nevertheless prove to be useful.

**Target Participants:** In our model validation study and survey study, we recruited participants ranging from early career researchers to expert researchers.

For the interview study, we targeted researchers with experience in teaching and providing research assistance to early career researchers, guiding them through the literature review process and the overall research journey. As Julien and Genuis [32] suggested, incorporating formal instructional opportunities into library and information studies programs and providing ongoing support within the workplace, educational institutions can ensure that librarians are well-prepared to teach information literacy. Experienced researchers and librarians, who are on the front line of teaching information literacy and are familiar with various models and frameworks, were particularly targeted. A more detailed description of interview participant recruitment

is provided in Chapter 4 (Section 4.3).

For the survey study, we targeted early career researchers who often face challenges and generally struggle in managing literature sources, conducting literature reviews, or creating literature databases [54]. A more detailed description of survey participant recruitment is provided in Chapter 5 (Section 5.4).

**Target Audience:** Here, we discuss the target audience who might benefit from our research findings or the application of our model. The findings from both the interview and survey studies helped to identify our target audience.

We hope that designers and developers who create tools and software for researchers to build their bibliographies will be among the target audiences and may find the results and discussion of this study useful. They may be able to gain insights into the needs of users, which may help them add new features or develop new tools to better meet these needs.

Early career researchers may be able to learn about the different characteristics of research groups and benefit from the proposed RaMSeS model, which outlines the general processes and phases an early career researcher goes through. They might reflect on their practices and the different stages of their research. The simulation of the RaMSeS model may help improve their mental model or the research literature review process. Additionally, the model may serve as a teaching tool for educators to teach research skills and methodologies to early career researchers, such as graduate students, and provide research assistance.

### 1.3 Organization of the thesis

The rest of the thesis is organized as the following chapters. In **Chapter 2**, we delve deeply into the background and related work, providing a comprehensive review of existing literature and theoretical frameworks that underpin our research. In **Chapter 3**, we present a detailed exposition of our proposed RaMSeS model.

**Chapter 4** outlines the methodology and findings of *interview study* which validates the RaMSeS model. **Chapter 5** outlines the methodology and findings of the survey study to capture the research habits of early-career researchers. This chapter covers the survey design, data collection methods, and analytical techniques. We present the results in a clear and systematic manner, highlighting key findings and their implications for our research questions. The rigorous methodological approach ensures the reliability and validity of our findings.

In **Chapter 6**, we discuss the results of both studies comprehensively and also offer recommendations for further investigation, suggesting specific areas where additional research is needed to build on our findings and enhance the impact of the RaMSeS model. Finally, **Chapter 7** concludes the thesis by summarizing the main findings and contributions of our research. We reflect on the significance of our work and its potential to influence future studies and applications.

## Chapter 2

# Background and Related Work

In this chapter, we provide an overview of the work in sensemaking and searching practices of knowledge practitioners and an overview of Reference Management Systems (RMS) and how knowledge practitioners employ them in their research process. Finally, we present the overview of Information Literacy (IL) Frameworks.

### 2.1 Information Seeking and Retrieval

The information seeking and retrieval behaviours of knowledge practitioners and researchers have been studied in the past, especially as the literature review process moved into the digital realm. Early work in this field, such as Hemminger et al. [28], surveyed information seeking practices of university academic scientific researchers. They found significant changes in information seeking behaviour due to the mass proliferation of online digital resources. They report increased reliance on web-based resources, fewer visits to the library, and almost entirely electronic communication of information. In a subsequent study, Niu et al. [50] surveyed researchers in five universities in the USA about their information seeking behaviours. They report similar information seeking behaviours in that the researchers' utilization of electronic methods for searching and accessing scholarly content dominated other mediums. These works have focused largely on the literature and reference gathering using online resources. However, more work needs to be done to understand how researchers conduct their overall research process [54].

Information seeking behaviours of researchers in specific domains have also been explored. For example, architectural students' behaviours to inform design [40], med-

ical researchers conducting systematic reviews [34], and biomedical researchers gaze, perception, and search behavior [38]. These works typically report useful findings to understand not only the information seeking behaviours performed in a particular domain but also the nature of the domain itself. For instance, Knight et al. [34] report that a system designed to support medical researchers to achieve the entire process (pre and post-search data extraction) would benefit from separating the processes. Athukorala et al. [5] report on computer scientists’ objectives and methods for doing literature searches. Their main findings suggest that respondents used search most frequently to remain current on issues. They also report on the difficulties encountered by researchers which involve exploring uncharted territory.

These works highlight the importance of understanding researchers’ information seeking practices, and the interest in quantifying and analyzing the difficulties that these researchers face. However, these works are limited in their applicability for general research usage as they are focused on particular domains and report findings to support researchers in the context of those domains. Our work takes a domain-agnostic approach to model and understand the overall research process. This is important as it can provide insights for developing generalizable tools for the overall research process.

These previously mentioned works have focused on the process of information seeking which leads to the collection of information repositories. Further work focuses on the systematic organization and retrieval of information from these complex collections [25]. Previous work reports that user concerns are crucial when designing information retrieval systems and having a Human-Computer Interaction informed perspective can lead to better information retrieval systems [25, 8, 63]. Our work adds to this field by conducting a user survey to understand the researchers’ perspective on the entire research process.

## 2.2 Sensemaking and Searching Practices of Knowledge Practitioners

Odden and Russ [51] consolidate fragmented definitions of “Sensemaking” in the context of science researchers as a process of generating explanations to resolve perceived gaps or conflicts in knowledge. Following these research directions, Marchionini [42] proposed that information seeking, sensemaking and learning are closely related.

Search as Learning (SAL) is defined as the process of constructing mental models of knowledge, retaining them over time, and applying them [67, 42]. SAL offers a broader perspective on search that extends beyond merely discovering and collecting information. However, most SAL research has focused on classroom students rather than knowledge practitioners like early career researchers [48]. According to Willoughby et al., [68], participants with higher domain expertise were more likely to experience the advantages of searching. Roy et al. [56] found that participants with greater prior knowledge gained new information, mostly toward a search session's end, while those with less prior knowledge gathered more new material early on in their search sessions. Chavula et al. [18] introduce SearchIdea, a web-based academic information search tool designed to enhance idea generation, organization, and synthesis of search results. Notably, university students using SearchIdea exhibit increased engagement and enhanced idea generation, particularly in terms of synthesizing and organizing ideas. Mysore et al. [48] delve into practices and challenges related to search and discovery, selection of search results, skimming, and reading among data scientists. They discovered that individuals face challenges when seeking and making sense of papers outside their disciplinary bubbles.

In our research, we captured the sensemaking and searching processes for the entire research process, specifically for knowledge practitioners with a relatively high level of prior knowledge (i.e. early career academic researchers) through the RaMSeS model (Chapter 3). We contribute by providing results from a survey to understand the search and sensemaking processes as they relate to the research process and how participants search, create and update their collection process.

## 2.3 Reference Management Systems (RMS)

Academic Reference Management Systems (RMS) serve as important tools for the research community by facilitating the organization, citation, and efficient communication of scholarly work [65]. Proske et al. [54] present information about 30 different RMS. Overall, they report that RMS streamlines creating bibliographies, collecting and organizing bibliographies, and facilitating collaboration with colleagues. Winslow et al. [70] stated that an RMS like Zotero can provide faculty and librarians with a platform that fosters productive, interdisciplinary collaboration. There has also been work to understand how specific techniques such as visualization can be leveraged into RMS to enable researchers to conduct collection building.

ResearchRabbit<sup>1</sup> [19, 43], enables users to discover publications related to one or more seed publications through the utilization of visualization maps and lists that encompass earlier, later, and similar publications. Mercadé et al. [44] conducted a study to investigate how network visualization tools, including ResearchRabbit, LitMaps<sup>2</sup>, and Connected Papers<sup>3</sup>, assist doctoral and undergraduate students in conducting literature reviews within real-world scenarios. However Proske et al. [54], mention that further research work needs to be done to understand the utilization of these tools, especially since researchers do not follow a linear process for collection creation and management. This is important in the context of our work as we also focus on modelling the entire research process, which is often not linear.

In addition, several works in the field of RMS have restricted themselves to focusing only on specific RMS or have demographically limited themselves to only studying the behaviours of researchers in specific domains or geographic locations. In a survey study, Basri et al. [6] gathered data from 100 participants from Indonesia to assess their responses regarding the utilization of Mendeley in the context of academic writing and reported that most students found Mendeley easy to use. Courraud et al. [20] surveyed medical writers and reported that they typically find Zotero to be a useful tool. Similarly, Kali [33] report RMS to be useful for medical literature management. Speare [58] surveyed students' RMS usage at the University of Manitoba and Nitsos et al. [49] reported a similar survey for students at Aristotle University of Thessaloniki.

These works offer important insights into the usage of such tools. Our work focuses instead on modelling the entire research process of early career academic researchers and not just the use of RMS, which constitutes an important component of that. In addition, in our survey study, we did not limit our work to be geographically or domain-specific, in an attempt to gain a fuller understanding of the research process of early career researchers. However, the interview study was limited in terms of the number of participants and that future work would require a larger number of participants and should include participants who work with researchers in other languages as well.

---

<sup>1</sup>[www.researchrabbit.ai](http://www.researchrabbit.ai)

<sup>2</sup>[www.litmaps.com](http://www.litmaps.com)

<sup>3</sup>[www.connectedpapers.com](http://www.connectedpapers.com)

## 2.4 Information Literacy (IL) Frameworks

Information literacy is a skill set for early career academic researchers, who must efficiently navigate, manage, and synthesize large volumes of information to construct knowledge and contribute to scholarly discourse. Framework for Information Literacy for Higher Education by the American Library Association [3] provides a set of core concepts and practices that outline the essential components of information literacy in an academic context. It emphasizes the importance of not only being able to locate and evaluate information but also understanding the ethical dimensions of information use and the ways information impacts knowledge production and dissemination.

The Big6 Information Literacy Model [24], by Eisenberg and Berkowitz, includes six stages: defining the information problem and identifying needed information, determining and selecting the best information sources, locating sources and finding information within them, engaging with information to extract relevant details, organizing and presenting information from multiple sources, and assessing the final product and the process used.

Bloom's Taxonomy [11], created by Benjamin Bloom and colleagues, is a hierarchical model that classifies educational learning objectives into six levels of complexity. These levels include recalling facts and basic concepts, explaining ideas or concepts, using information in new situations, breaking down information into parts to explore relationships, justifying decisions or actions, and generating new ideas or products. Each level builds on the previous one, requiring increasingly complex cognitive processing. The taxonomy provides a framework for educators to design curricula and assessments that foster higher-order thinking skills, promoting critical thinking and problem-solving abilities in students.

The concept of "Undiscovered Public Knowledge," introduced by Don R. Swanson highlights the potential for discovering new knowledge through the synthesis of existing research across different disciplines. He argued that valuable information is often scattered across various fields, and connecting these pieces can reveal novel relationships and generate new hypotheses. It emphasizes the value of comprehensive literature reviews, cross-disciplinary research, and advanced information systems in facilitating the discovery process, encouraging scholars to explore the broader landscape of available information for new insights [59].

**Understanding and Enhancing Information Literacy:** Information literacy extends beyond the ability to search for information; it includes the critical evaluation

and organization of information, synthesis of ideas, and creation of new knowledge. Lloyd defines information literacy specifically within the context of academic research as a “culture of practice,” not merely a set of skills but an integral part of scholarly practice that evolves with experience and interaction within a community [39].

Moreover, Julien and Genuis argue for the need to integrate information literacy into the core curriculum of higher education, suggesting that this integration can significantly enhance the research output of new academics by providing them with better tools and strategies for managing information [32].

Eisenberg covers conceptual understandings of IL, IL standards, models, and technology within the IL framework [23]. He compared various models of IL that were developed through research, practice, and committee, respectively. His side-by-side view of IL models shows that there are many similarities among them. He also mentioned the Big6 a process model of how people of all ages solve an information problem.

**The Role of Technology in Information Literacy:** Digital tools and Reference Management Systems have transformed how researchers manage their literature collections. A study by Isfandyari-Moghaddam et al. explored the role of information technology in IL which recommends mandatory lessons related to research methods and IL to students [31].

Looking forward, there is a growing recognition of the need to further integrate information literacy training into doctoral education programs. Ince et al. mentioned four areas of training (information literacy, information management, knowledge management and understanding the lifecycle of scholarly communication) to be considered by librarians and faculty advisers to better prepare scholars for their future [30].

These works provide a detailed foundation for understanding the critical role of information literacy how models like RaMSeS may impact the lives of early-career researchers and the design of new tools and features for researchers. Our work offers an information literacy framework, characterizes early-career researchers’ behaviour, and validates the proposed model, highlighting the importance and applicability of an IL model like RaMSeS.

## Chapter 3

# Researcher's Management of Search and Storage (RaMSeS) Model

The act of searching through databases is known to be laborious and frustrating, regardless of whether one is a novice or an experienced user [69, 41, 29]. Given the ever-increasing nature of academic literature, researchers have proposed several information practice models to guide the process of sensemaking. Such models include the classical Information Retrieval (IR) model [9], the IR model augmented for the web [14], and bookmarking indexing terms [53]. Blake et al. [10] conducted an observational study of academic research scientists in Collaborative Information Synthesis and proposed the Collaborative Information Synthesis (CIS) Model. These models explored various aspects of a researcher's journey such as online searching [53] and collaborative work [10].

While previous works focused on specific aspects of the research process, our proposed **R**esearcher's **M**anagement of **S**earch and **S**torage (RaMSeS) model simulates the entire academic research process, offering a comprehensive framework for managing search and storage activities. This model describes the research process in a way that can be particularly valuable as an educational aid. In Chapter 4, we investigated the effectiveness of the RaMSeS model with experts in academic literature research. They found that it can serve as a teaching tool to better explain the research process and support students in developing effective research strategies.

Inspired by the insights garnered from the previously introduced models, and the

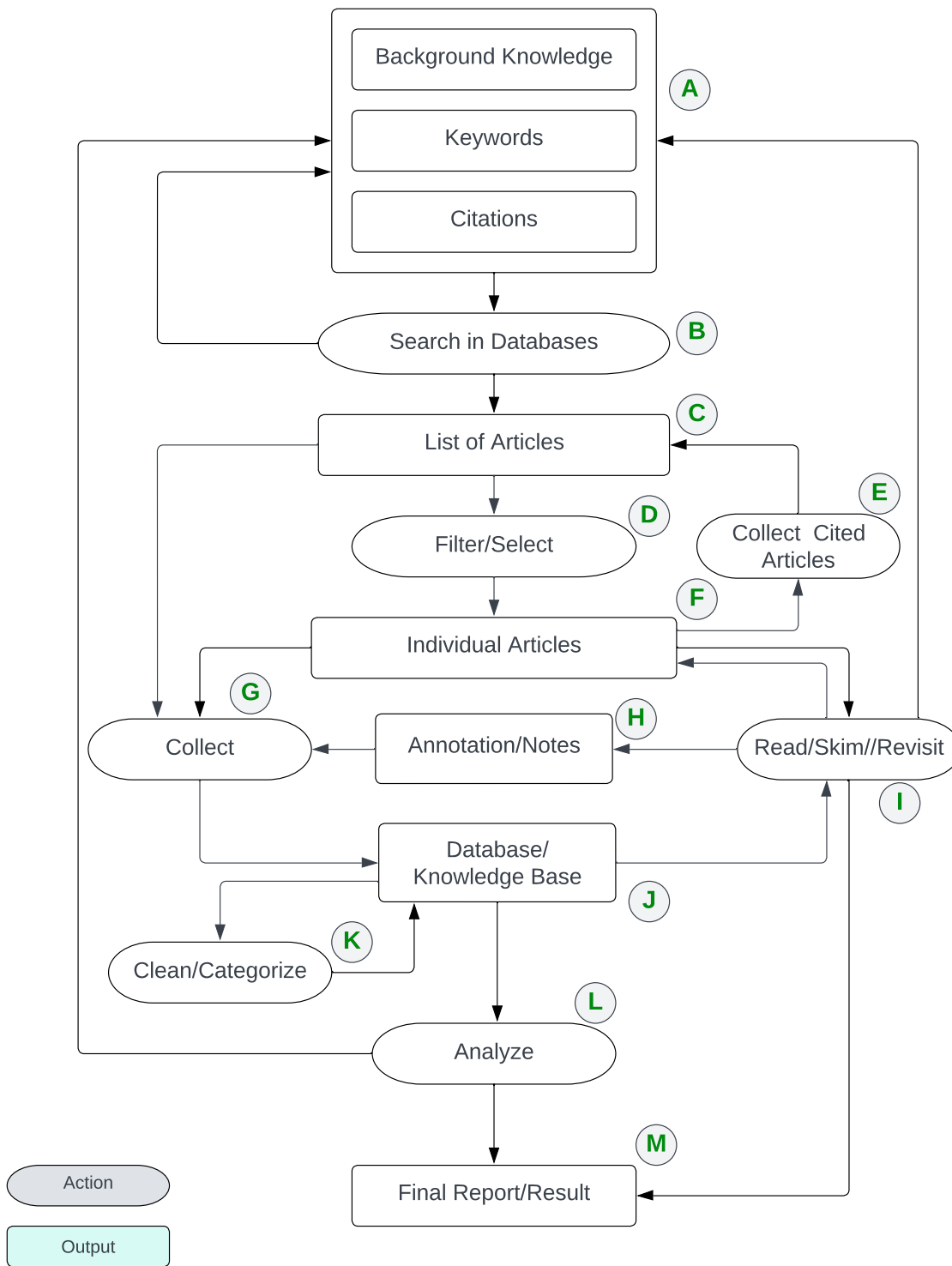


Figure 3.1: Researcher's Management of Search and Storage (RaMSeS) Model

model comparison in subsection 3.3, we iterated over several model designs to devise the RaMSeS model (Figure 3.1), tailored specifically for academic researchers. We modelled the research process in actions and their outputs. An action typically results in an output, which can then be the input for another action. For example, some researchers may loop several times in collecting cited articles from a selected paper, (Figure 3.1.E) and some may not take notes at all (Figure 3.1.H).

The list of actions and outputs:

- Actions
  - Search in Databases (B)
  - Filter/Select (D)
  - Collect Cited Articles (E)
  - Collect (G)
  - Read/Skim/Revisit (I)
  - Clean/Categorize (K)
  - Analyze (L)
  
- Outputs
  - Background Knowledge, Keywords, Citations (A)
  - List of Articles (C)
  - Individual Articles (F)
  - Annotation/Notes (H)
  - Database/Knowledge Base (J)
  - Final Report/Result (M)

To evaluate its practicality and relevance, we held an informal meeting involving early career and advanced academic researchers at our weekly lab meeting. During this meeting, the lab members were provided with a printed copy of the RaMSeS model and were invited to simulate its application in their research process. We encouraged them to assess how well their information practices aligned with RaMSeS and to provide feedback by annotating their observations.

The information seeking typically commences with background knowledge, a keyword search, and citations (Figure 3.1.A) in online resources/databases like Google

Scholar (Figure 3.1.B). From the search results obtained (Figure 3.1.C), researchers engage in the subsequent tasks of filtering, selecting, and collecting relevant sources (Figure 3.1.D). This selection process results in the compilation of a new list of related works (Figure 3.1.E), thereby enriching researchers' background knowledge and refining their selection of keywords. Each individual article (Figure 3.1.F) is subject to various levels of examination, which may include comprehensive reading, skimming, or collection for future reference (Figure 3.1.I). Reading these articles often leads to note-taking (Figure 3.1.H), with the information being organized and stored (Figure 3.1.G) in a database for future retrieval and reference (Figure 3.1.J). The collected materials often undergo a curation process, during which they are cleaned, organized, and categorized (Figure 3.1.K) to ensure their relevance and to maintain an up-to-date repository. Ultimately, all the collected resources are subjected to analysis (Figure 3.1.L), for the composition of the final research report (Figure 3.1.M).

### 3.1 Description of RaMSeS Terms

Our proposed RaMSeS model consists of various terms as *actions* and *outputs* within its framework (e.g., searching in databases, collecting, note-taking, reading and skimming, revisiting, analyzing and writing). We are describing the terms below:

*Search in Database:* This is one of the most common operations in the literature research domain, and it has received much attention. Researchers query a database to find documents that are relevant to their objectives. Most often it involves typing keywords in a search box of a database such as Google Scholar, but searches might be arbitrarily complex, sometimes depending on the database. For example, with some databases, one could search by a combination of different attributes, such as author, keywords, or research domain. Searches can also involve logical operators. The output of searches is usually a list of one or more articles that comply with the search term.

*Collect:* In our context, *collect* refers to the act of gathering materials and information in a more permanent form with some sort of structure. For example, a researcher might read a set of articles on paper and keep several folders of those articles organized by topic. They may save source files in different formats such as URLs, images, videos, documents, and more, organizing them into collections within their storage systems and knowledge bases. This organization can take various forms, including the utilization of different reference management tools [73, 15], the book-

marking of web pages within web browsers, or even the incorporation of these files into software systems like Microsoft Word, Excel, and other note-taking applications. Furthermore, the collection and organization of written information were reviewed in an exploratory study by Case [17].

*Database and Knowledgebase:* In our work database and knowledgebase *refers* to some permanent or semi-permanent storage of information regarding the sources, searches, keywords, and documents that have been involved in the process of literature research. Databases can take forms from the very simple (e.g., a collection of citations in a plain text document) to very sophisticated (a searchable repository of documents that have been collected or read, with notes). Some people might also not produce permanent storage of information and instead keep this kind of information in their memory, which is why we refer also to the knowledge base.

*Read and Skim:* "Read" and "skim" are two different reading techniques used to process written material, such as books, articles, or documents, with varying levels of depth and attention. Adler and Doren published a manual on intelligent reading [1], while Rocha and Nacenta conducted an exploratory study on the impact of environment and technology on people's ability to read well in academic settings [46]. In another study comparing reading and skimming from computer screens and books, it was found that skimming was 41% slower when performed on CRTs compared to using traditional printed books [47].

*Note-taking:* Notes can be taken in diverse forms, like handwritten or digital notes, outlines, diagrams, and annotations, depending on preferences and usage context. In our work, we refer to note-taking which can be commonly done by researchers during or after reading or skimming. Sönke Ahrens introduced an effective note-taking methodology to stimulate new ideas on paper [2]. Leong discusses the benefits of using different types of notes, including bibliographic notes, summary notes, and analytic notes, to capture different kinds of information [37].

*Revisit:* Revisiting involves the process of reading sources and notes previously collected. The collected data and information may be revisited periodically for communication or final report writing. Toba et al. investigate users' revisitation behavior based on web access [64].

*Analyze:* The analysis of collected information and notes involves a comprehensive interpretation of gathered data to uncover patterns, insights, and key findings. This process is crucial for gaining a deeper understanding of the data's significance, which is essential for final report writing purposes. Meta-analysis, on the other hand, is

defined as the analysis of a large collection of analysis results with the purpose of integrating the findings [27, 22].

*Writing:* Writing is more than just the final stage after analysis; it is integral throughout the entire research process. It functions as a social technology for communication, involving meaning, goals, and identities within a dynamic social context [7]. Writing happens not only in the final synthesis but also during activities like annotating and writing to learn.

## 3.2 Alice’s Journey through the RaMSeS Model

To provide a better understanding of the model in a storytelling manner, we narrate how Alice takes *actions* (highlighted in red) and the resultant *outputs* (highlighted in green), as referenced in Figure 3.1.

Alice is an early career researcher who has just started her PhD in Biochemistry. She has some background knowledge of her research area and has come across related work (citations) that are most relevant to her, learning keywords from those papers Fig 3.1.A).

Alice is conducting a literature review and performing literature searches (Fig 3.1.B) across different scholarly databases like Google Scholar<sup>1</sup> and PubMed<sup>2</sup>. Alice comes across a list of articles (Fig 3.1.C) from different searches across these databases. She meticulously filters and selects (Fig 3.1.D) the most relevant articles from these lists. As Alice filters search results and goes through individual articles (Fig 3.1.F), she collects cited articles (Fig 3.1.E) that appear relevant to her, thereby building a more comprehensive collection of related works. Each article Alice selects undergoes various levels of examination. She reads some in detail, skims through others, and sets aside a few for future reference.

Alice collects (Fig 3.1.G) these individual articles and organizes them for later use. Sometimes she takes notes and annotates (Fig 3.1.H) key points as she reads through the articles. The process of reading, skimming, and revisiting articles (Fig 3.1.I) becomes an iterative task for Alice, allowing her to refine her understanding continually. Alice stores her selected articles, and notes in a reference management system/database called Zotero for future retrieval and reference (Fig 3.1.J). She cleans and categorizes (Fig 3.1.K) her stored materials to ensure that the repository re-

---

<sup>1</sup><https://scholar.google.com>

<sup>2</sup><https://pubmed.ncbi.nlm.nih.gov>

mains relevant and up-to-date. She then begins to analyze (Fig 3.1.L) the collected resources to draw insights and support her research findings. Ultimately, all the collected resources and analyses culminate in the final research report (Fig 3.1.M) for Alice.

Throughout Alice’s journey, we can see that in her literature search to write the final report the actions and outputs of the RaMSeS model are reflected clearly.

### 3.3 Comparison of RaMSeS to other models

The RaMSeS model is designed to address the entire academic research process, distinguishing itself from other models by simulating the sequence of actions a researcher undertakes from initial search to final analysis. In this section, we have done a comparison of RaMSeS to other models. Table 3.1 provides a summary of how RaMSeS aligns with and differs from the three most recent prominent models.

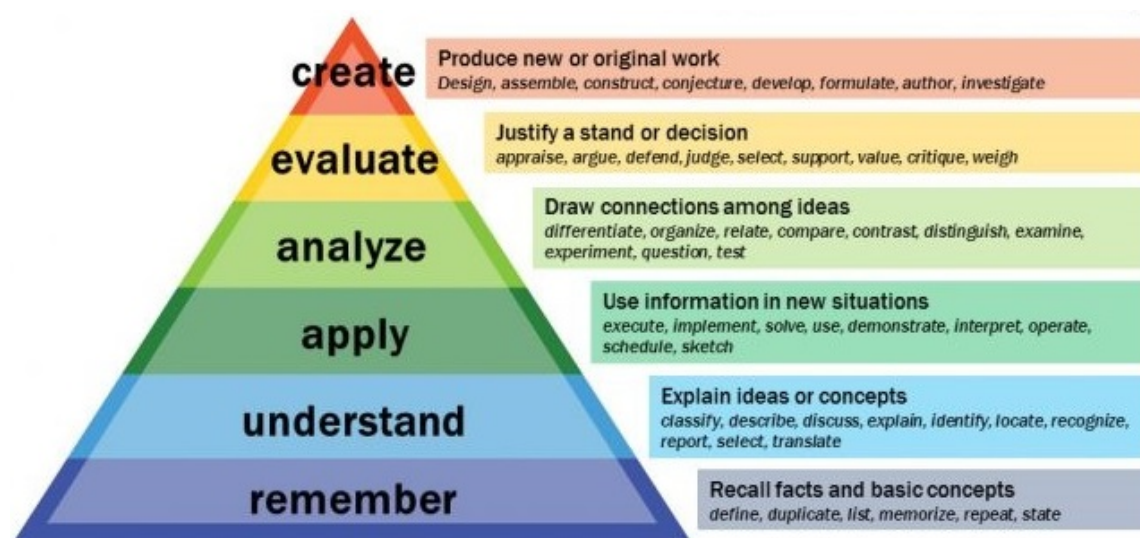


Figure 3.2: Bloom’s Taxonomy [4]

**Bloom’s Taxonomy:** Developed by Benjamin Bloom and his colleagues in 1956, Bloom’s Taxonomy is a hierarchical model used to classify educational learning objectives into levels of complexity and specificity: Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating. Each level builds on the previous one, requiring a higher level of cognitive processing [11]. This taxonomy provides a framework for educators to design curriculum, assessments, and instructional methods that

foster higher-order thinking skills. It encourages students to move beyond memorization to deeper levels of understanding and application. Bloom's Taxonomy provides a hierarchical framework, while RaMSeS focuses on the practical steps involved in academic research. RaMSeS aligns with Bloom's higher-order thinking skills, particularly in the analyzing and creating stages, by providing a structured approach to synthesizing information and generating new knowledge. RaMSeS's systematic process may help researchers achieve deeper levels of understanding and application, as outlined in Bloom's Taxonomy.



Figure 3.3: The Big6 Information Literacy Model [26]

**The Big6 Information Literacy Model:** Created by Eisenberg and Berkowitz,

the Big6 is a structured framework designed to teach and apply information literacy skills through six stages: Task Definition, Seeking Strategies, Location and Access, Use of Information, Synthesis, and Evaluation [24]. The Big6 guides students through a comprehensive problem-solving and research process, promoting critical thinking, effective information management, and lifelong learning skills. Big6 focuses on teaching and applying information literacy skills in an educational setting. RaMSeS, on the other hand, is more oriented towards the specific needs of academic researchers, detailing actions and outputs to manage the research process effectively. RaMSeS can be seen as a more specialized extension of the Big6, tailored to the research context.

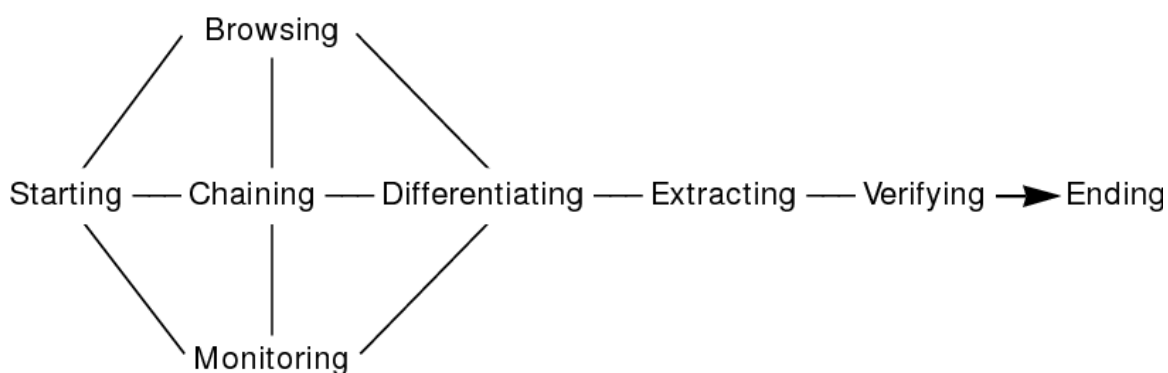


Figure 3.4: Ellis’s Model of information seeking Behavior [69]

**Ellis’s Model of information seeking Behavior:** Ellis identifies specific behaviors such as starting, chaining, and browsing [25]. RaMSeS provides a systematic approach to these behaviors, particularly in how researchers filter, select, and organize information. For example, the process of filtering and selecting articles (Fig 3.1.D), collecting cited articles (Fig 3.1.E), and organizing them into a database (Fig 3.1.J) mirrors Ellis’s behaviours of differentiating, chaining, and monitoring. RaMSeS offers a more granular roadmap for academic research by breaking down these behaviors into specific actions and outputs, thus guiding researchers through the detailed steps necessary to effectively gather and manage information.

**Kuhlthau’s Information Search Process (ISP):** Kuhlthau outlines a six-stage process with emotional and cognitive aspects. RaMSeS aligns with Kuhlthau’s ISP by detailing the practical steps of research, offering a tangible structure to the somewhat abstract stages of Kuhlthau’s model. For example, RaMSeS provides specific actions such as searching in databases (Fig 3.1.B), filtering/selecting articles (Fig 3.1.D), and annotating notes (Fig 3.1.H), which correspond to the selection, exploration, and formulation stages of Kuhlthau’s model [35]. This alignment makes it

Table 3.1: Comparison of RaMSeS with Other Models

Model	Description	Comparison to the RaMSeS Model
<b>Kuhlthau's Information Search Process (1991)</b>	Outlines a six-stage process with emotional and cognitive aspects.	RaMSeS aligns with Kuhlthau's ISP by detailing practical research steps corresponding to ISP's stages, helping researchers navigate emotional and cognitive challenges with clear, actionable steps.
<b>Wilson's Model of Information Behaviour (1999)</b>	Focuses on the contextual and psychological barriers to information seeking, shifting emphasis from utilization of sources to use of information by people.	RaMSeS reduces barriers by offering a structured research framework, organizing and streamlining search and storage phases, and addressing personal and environmental factors with systematic processes.
<b>ACRL Framework for Information Literacy (2015)</b>	Consists of six interconnected frames for students to navigate the information landscape effectively, engage in scholarly discussions, and use information ethically.	RaMSeS complements ACRL by offering practical steps for managing research activities and enhancing strategic exploration and inquiry aspects of information literacy.

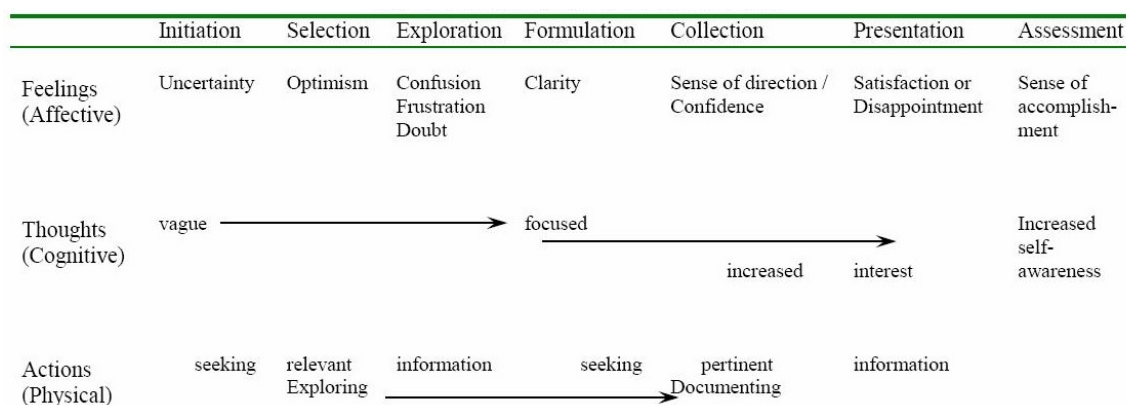


Figure 3.5: Kuhlthau's Information Search Process [57]

easier for researchers to navigate the emotional and cognitive challenges by following clear, actionable steps.

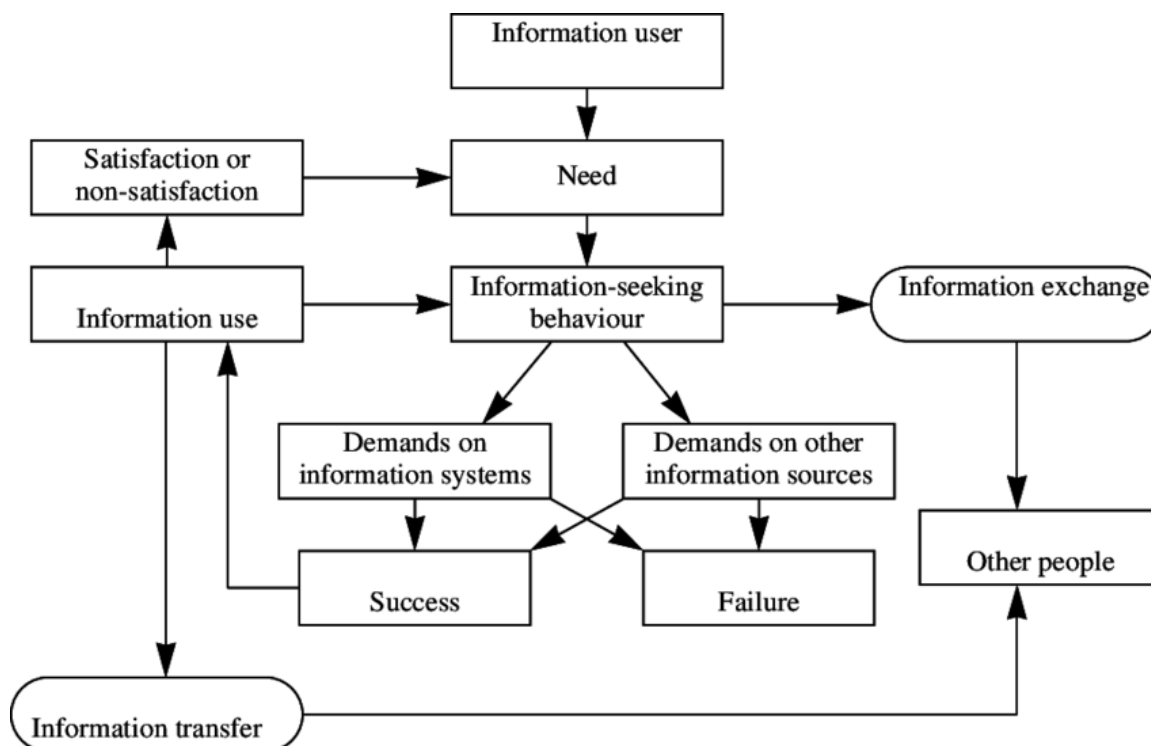


Figure 3.6: Wilson's Model of Information Behaviour [69]

**Wilson's Model of Information Behaviour:** Wilson developed his information seeking behaviour model in response to the need to shift the emphasis of information and library science from the utilization of sources to the use of information by people. Wilson's model focuses on the contextual and psychological barriers to information seeking. The RaMSeS model complements Wilson's approach by offering a structured framework to navigate the research process, thereby reducing these barriers. RaMSeS aids in organizing and streamlining the search and storage phases of research. It addresses personal factors through background knowledge and note-taking, supports interpersonal factors with collaborative tools, mitigates environmental barriers with systematic search processes, and overcomes systemic issues by efficiently filtering and collecting relevant information [69].

**ACRL Framework for Information Literacy:** The ACRL framework, developed by the Association of College and Research Libraries, consists of six interconnected frames: Authority Is Constructed and Contextual, Information Creation as a

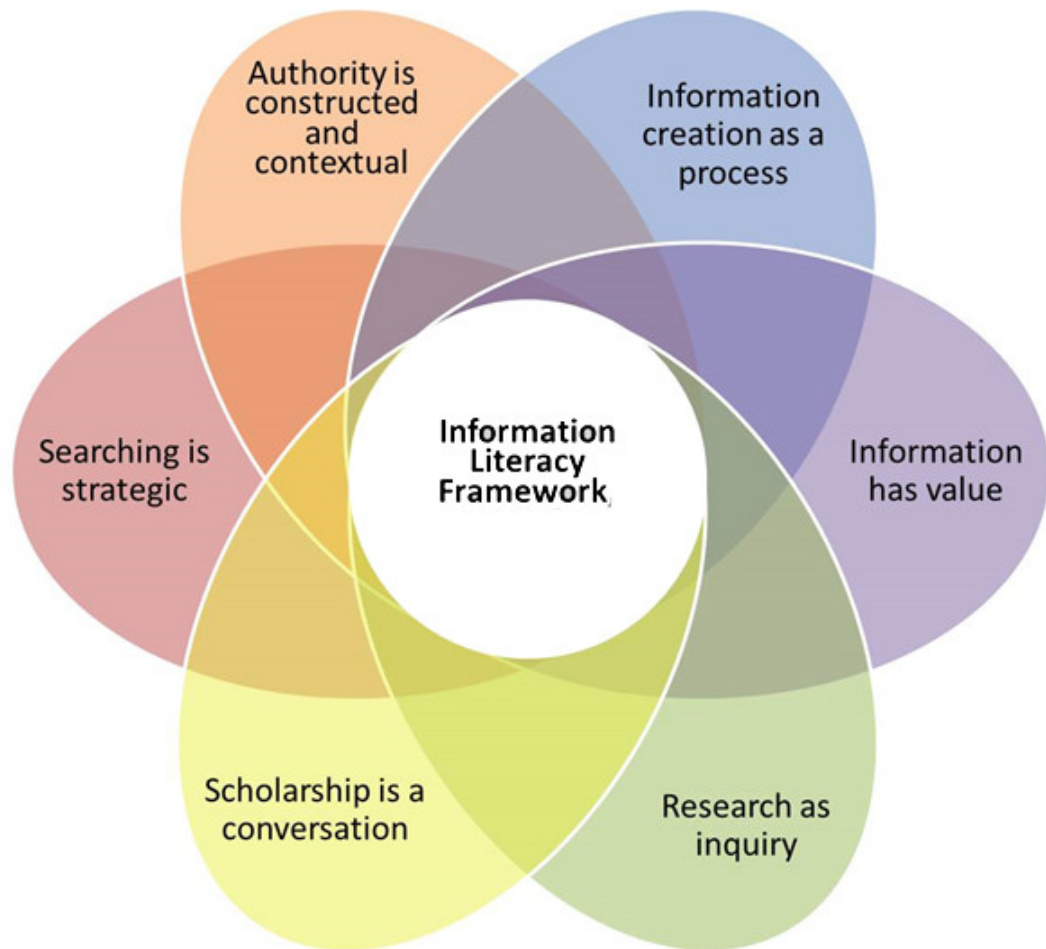


Figure 3.7: The ACRL Framework for Information Literacy [66]

Process, Information Has Value, Research as Inquiry, Scholarship as Conversation, and Searching as Strategic Exploration [3]. This framework aims to develop critical thinking and reflective skills, enabling students to navigate the information landscape effectively, engage in scholarly discussions, and use information ethically. The ACRL Framework focuses on the broader conceptual understanding and ethical use of information, while RaMSeS offers a practical step-by-step guide for managing research activities. RaMSeS complements the ACRL Framework by providing detailed actions and outputs for each stage of the research process, enhancing the strategic exploration and inquiry aspects of information literacy.

Overall we can see that the RaMSeS's usefulness lies in its step-by-step simulation of the academic research process, incorporating aspects of these models while focusing on the practicalities of research management. It stands out by not only addressing the behavioral and cognitive aspects of information seeking but also integrating these into a coherent, actionable framework that spans the entire academic research lifecycle. This practical emphasis on the sequence of traditional academic research activities, from initial searching to the organization and analysis of information, marks a significant contribution to the field, bridging theoretical models with practical application.

## Chapter 4

# Validation of the RaMSeS Model

The RaMSeS model that we introduced in Chapter 3 offers a lens to the processes and documents that take place in research activities related to managing literature and sources as well as writing the types of outputs that rely on that data. Although useful for our thinking processes and the formulation of research questions, this model could potentially have blind spots, biases, or flaws that we could not see ourselves. As a form of validation of the model we carried out a study with experts in the domain that could expose potential problems and that allows us to verify whether these experts find value in thinking about the research activity through the lens of RaMSeS. The study took the shape of a semi-structured interview.

This chapter describes the design of our Semi-structured interview study and participant sample, study procedure, and findings. To maintain consistency and clarity, we positioned this interview study (in Chapter 4) after the introduction of the model (in Chapter 3).

### 4.1 Interview Research Questions

The interviews aimed to validate the RaMSeS model with feedback from experts in this field. We used a qualitative approach to have deep discussions with these experts about how they conduct literature reviews, learning about their methods and experiences. This approach was chosen because it allows us to explore individual viewpoints and processes in detail, providing rich information that we might miss with other methods. By having open-ended conversations, we could understand how experts do their literature reviews and then see if the RaMSeS model fits and supports

their practices.

- RQ1 - Does the RaMSeS model comprehensively represent the key activities and sequences undertaken by literature reviewers in academic research?
- RQ2 - To what extent do experts in research literature processes find the RaMSeS model descriptive and useful for developing their research methodologies?
- RQ3 - Is the RaMSeS model useful for describing normative or descriptive processes in literature research theories and current practices?
- RQ4 - What are the primary limitations or missing aspects in the RaMSeS model as identified by experts in research literature processes?

## 4.2 Interview Design

We designed the semi-structured interview inspired by insights and findings from the survey study (Chapter 5) we conducted before this interview study. To ensure continuity in this thesis we positioned the interview study after the introduction to our proposed model.

Interview plans typically include a range of questions and prompts crafted to direct the interview in a focused yet adaptable and conversational manner [45]. From our research questions (Section 4.1), we developed a series of interview questions (see Table 4.1, and Appendix A) that underwent iterative refinement across multiple phases. These questions are grouped into two parts: the first part (Table 4.1, Phase 2A) explores participants' research processes, while the second part (Table 4.1, Phase 2B) assesses how well the model aligns with participants' methods and strategies, offering valuable insights into its effectiveness.

## 4.3 Participant Recruitment

We employed a systematic recruitment strategy to identify suitable participants for our study. Following the approval of our harmonized ethics application, recruitment advertisements were circulated via email to universities listed in the ethics protocol. Prospective participants were required to meet specific criteria: they had to be researchers, academics, professors, or librarians with expertise in managing research literature processes. Interested individuals were informed that participation

involved a screening process, reading and signing a consent form, completing the interview study, and optionally filling out a post-study demographics survey administered through our institution-licensed Survey Monkey<sup>1</sup>.

We successfully recruited five participants from various universities across British Columbia, Canada. This diverse group included four experienced librarians specializing in research literature management and providing research assistance, alongside a professor from a technical writing program. Two participants represented the arts/humanities domain, one the sciences, one the social sciences, and another defined their domain as Academic Librarian, specializing in Open Access and GenAI.

Participants were offered CAD\$15 as compensation for their involvement, and all of them waived the compensation. The study proceeded according to the approved experiment protocol (see Table 4.1, and Appendix A) by initiating a consent and introduction phase (Phase 1), during which participants reviewed and signed the consent form received via email. The main interview, spanning 30 to 40 minutes, was structured into three phases (2A, 2B, and 2C). Final Phase 3 concluded the study with a debrief.

## 4.4 Study Procedure

The study procedure involved several phases, ensuring comprehensive data collection and participant engagement. Initially, the interview began with a consent and introduction (*Phase 1*), lasting approximately one minute, as participants were provided with the consent form via email, which they reviewed and signed before the actual interview.

The main interview segment spanned 30 to 40 minutes and was divided into three distinct *Phases 2A, 2B, and 2C*. The first 15 minutes (Phase 2A) focused on capturing participants' research processes. To avoid biases towards our RaMSeS model, we didn't explain or show our model in Phase 2A. Instead, we focused on capturing participants' methods for conducting literature reviews and assisting students in their research. Participants were encouraged to illustrate and describe their approaches using a digital space called Microsoft Whiteboard<sup>2</sup>. We provided the whiteboard link via the meeting chat. However, none of them used the whiteboard in Phase 2A. They were prompted with some questions (see Table 4.1) about the phases of the process,

---

<sup>1</sup>[www.surveymonkey.com](http://www.surveymonkey.com)

<sup>2</sup><https://app.whiteboard.microsoft.com>

the order in which they occur, the areas they focus on the most, the challenges faced, and the tools recommended for organizing and managing their literature findings

Table 4.1: Interview Plan and Questions

Phases	Interview plan and Questions
Phase 1	Consent and Introduction (1 Minute)
Phase 2A	<p><i>Activity to Capture participants research processes (15 Minutes)</i></p> <ol style="list-style-type: none"> <li>1. Can you first explain the different phases in the entire process, from ideation to writing the related work section?</li> <li>2. Do these phases you mentioned always occur in the same order?</li> <li>3. Which part of the process do you FOCUS on the most?</li> <li>4. Which parts do students or researchers handle effectively?</li> <li>5. Which aspects of the process are MOST challenging?</li> <li>6. Is the description you've given me related to a specific model or ontology that you've encountered elsewhere?</li> <li>7. What techniques or tools do you recommend to students for organizing and managing their literature findings?</li> </ol>
Phase 2B	Introduction to RaMSeS Model (5 Minutes)
Phase 2C	<p><i>Interactive Session Using MS Whiteboard (15 Minutes)</i></p> <ol style="list-style-type: none"> <li>8. How closely does our model mirror the methods and strategies you've explained earlier?</li> <li>9. Does our model accurately capture all aspects of your process, or are there elements included that do not reflect your actual processes?</li> <li>10. How do you think a structured model like RaMSeS could influence or enhance the way you teach research methodology and processes to students?</li> <li>11. Based on your experience, what improvements or additions could make the RaMSeS model more effective as a teaching tool for the research process?</li> <li>12. What are your overall thoughts on this model?</li> </ol>
Phase 3	Conclusion and Debrief (1 Minute)

Following this, a five-minute introduction (*Phase 2B*) to the RaMSeS model. Participants were walked through the model and encouraged to ask questions and provide feedback. The next 15 minutes (*Phase 2C*) involved an interactive session using the MS Whiteboard. Participants were asked to compare the model with their methods and strategies and to provide insights on how accurately the model captures their processes. Feedback was gathered on the potential impact of the RaMSeS model on teaching research methodology and any improvements or additions that could enhance its effectiveness as a teaching tool.

The session concluded with a debrief and conclusion phase, lasting about one minute. Additionally, a survey link was sent over email after the interview to collect optional demographic information. The detailed interview questions and plan can be accessed in the Appendix A.

## 4.5 Data Collection and Analysis

Our data collection involved video recordings and transcripts obtained via our institution-licensed MS Teams<sup>3</sup> throughout the interview. To analyze the qualitative data, we examined the transcripts of the interviews. The transcripts were cleaned, anonymized, and then coded using institution-licensed NVivo<sup>4</sup> by the primary researcher. The review of the codebook was conducted by a three-member research team.

Qualitative data analysis delves into both the unique themes of meaning found within the data and the common themes that arise across various data points [62]. Thematic analysis may involve several phases: familiarizing with the data, generating codes, constructing themes, reviewing potential themes, defining and naming themes, and producing the report [61]. In our thematic analysis, we employed most of the phases discussed by Terry et al.[61], ensuring a systematic and thorough examination of our data.

The final version of the codebook (Table 4.2), which consists of the codes derived from the data, was utilized for thematic analysis. This process led to the identification of key themes that provided answers to the research questions along with insights related to the interview questions. The detailed codebook used for the analysis can be accessed in the Appendix A.

## 4.6 Interview Findings

The findings from the thematic analysis are organized in terms of the interview phases. In Phase 2A, we asked participants about their own literature process, with their findings described in sub-section 4.6.1 to 4.6.6. Participants discussed their literature review phases, including the order, the most focused phases, the most challenging phases, and their awareness of any models they follow or know.

In Phase 2B, we introduced our proposed RaMSeS model and asked follow-up questions in Phase 2C. Findings are described in subsections 4.6.7 to 4.6.10. Participants discussed the representation and applicability of the RaMSeS Model, its usefulness in teaching research methodology, suggested improvements for enhancing RaMSeS as a teaching tool, and techniques and tools for organizing and managing literature findings. We are providing the codebook here in Table 4.2, also a detailed

---

<sup>3</sup><https://www.microsoft.com/en-ca/microsoft-teams>

<sup>4</sup><https://lumivero.com/products/nvivo>

version of this codebook is available in Appendix A.

Table 4.2: Interview Codebook

Code Name	Files	References
<b>01-05. Participants' Literature Review Process</b>	<b>5</b>	<b>88</b>
Literature Phases	5	60
Phases Order	2	3
Focus Areas	4	5
Effective Handling	5	5
Challenges	5	9
<b>01.02. Research Tools and Techniques</b>	<b>5</b>	<b>48</b>
Background Knowledge	3	9
Databases	4	10
Filtering and Selection	1	2
Keywords	4	8
Research Questions	4	5
Search Techniques	5	23
<b>06. Models and Frameworks</b>	<b>5</b>	<b>19</b>
ACRL Framework for Information Literacy	3	6
Anne Lamott's concept of 'shitty first drafts'	1	1
Big6 Skills of Information Literacy	1	1
Bloom's taxonomy	1	1
Personal Approach from Experience	3	4
The Metamorphosis by Franz Kafka	1	1
<b>07.01. Tools Recommendation in Literature Management</b>	<b>5</b>	<b>20</b>
Citation Managers - EndNote, Mendeley, or Zotero	3	5
Non-Users of Citation Managers	1	1
Pen-and-Paper	1	2
Simple Tools - Word or Excel	2	2
<b>07.02. Teaching</b>	<b>4</b>	<b>24</b>
Citation Management	3	3

Generative AI	1	3
Method Teaching	4	9
Research Management	4	6
Sensemaking	3	5
<b>08-09. Evaluation of RaMSeS Model</b>	<b>5</b>	<b>8</b>
Close Representation	3	12
Comprehensive Model	2	5
Thorough Model	1	3
<b>10. RaMSeS Enhancing Teaching Research Methodologies</b>	<b>5</b>	<b>11</b>
Core Emphasis	1	2
Helpful Guide	4	8
Overarching Model	3	5
<b>11. RaMSeS Scope of Improvements</b>	<b>5</b>	<b>26</b>
Gen AI capabilities	1	4
Improvement suggestions	4	18
Layout to a story	1	2
Structured searching	1	3

#### 4.6.1 Literature Review Phases

Participants described their research process, before exposure to our model, which follows a nearly linear logic, beginning with *Ideation*.

**Ideation:** This phase involves brainstorming and mind mapping. This step is crucial as it sets the foundation for the research by refining the focus and ensuring the research question is clear and specific. P2 said, “*Ideation often occurs when they come to us as librarians; they have already been through the phase where they have to come up with a research topic and question.*” P3 further emphasized the importance of narrowing down the topic to a manageable scope by asking targeted questions, which helps in pinpointing the exact focus of the research. This illustrates that ideation is an iterative process where researchers continuously refine their questions and concepts before proceeding.

**Searching:** Following ideation, the Searching phase involves selecting appropriate tools and strategies to gather relevant literature. P1 recommended starting with

a discovery layer that pre-indexes all the library's databases, emphasizing the importance of Boolean searches and advanced search features to refine search results. P2 and P4 underscored the necessity of using multiple databases to gain a comprehensive view of the literature. P2 stated *"We often start with the library catalog, which is kind of an interdisciplinary academic database. Once we've identified a few databases. I usually suggest not only looking at one but trying at least two to get a more complete picture."* P4 added, *"When starting a literature review, I typically begin with Google Scholar to find articles of interest. Sometimes, I also use Scopus, especially for topics that are more scientific or technical,"*. These strategies ensure a thorough and well-rounded search, capturing a wide range of relevant sources.

**Organizing:** Once the relevant literature is gathered, the Organizing phase comes into play. This phase involves managing and categorizing the collected literature for easy retrieval and reference. P1 elaborated on using citation managers like Zotero<sup>5</sup> and Mendeley<sup>6</sup> to create folders and subfolders, organizing articles into broad classifications. They stressed the importance of continuously categorizing articles as less or more essential and using digital highlights to keep track of important information. P5 highlighted the significance of collecting resources generously during the search phase, stating, *"I emphasize collecting generously as you go so that if something turns out to be relevant, you have it in your personal library."* This systematic organization helps maintain a structured approach, making it easier to manage a large volume of sources.

**Final Report Writing:** The final phase is more than just the process of writing the literature review and related work section. It involves synthesizing the collected information and presenting it coherently. Writing occurs throughout the research process, not just at the end. For instance, annotating sources and drafting preliminary notes are integral to shaping the final document. P1 noted that when conducting research for themselves, they could skip the preliminary steps and focus directly on writing the research question, selecting the appropriate database, and diving into the literature. P3 discussed teaching students how to summarize articles, paraphrase effectively, and avoid issues like patchwriting and plagiarism. P3 explained, *"After they understand summarizing and annotated bibliographies, we focus on supporting their arguments and being persuasive in their writing."* This final writing phase is essential for integrating literature into a coherent narrative that supports the research ques-

---

<sup>5</sup><https://www.zotero.org>

<sup>6</sup><https://www.mendeley.com>

tion and provides a robust foundation for the study, reflecting how writing contributes throughout the research journey.

Overall, the literature review process described by participants follows a structured yet iterative approach encompassing ideation, searching, organizing, and writing. Each phase is crucial in building a comprehensive and coherent literature review that effectively supports the research question.

#### **4.6.2 The Dynamic Nature of the Research Process**

Participants also emphasized that the research process is dynamic, with researchers often moving back and forth between different phases. This flexibility allows for the exploration of new ideas and the incorporation of new information. P1 described this non-linear nature, stating, *“It’s never a straight line, though it starts broad, becomes more specific, and might expand again as needed. It’s more of a winding path than a circle.”* This highlights how researchers begin with broad searches, and then narrow down to specific topics, demonstrating the adaptable and iterative nature of the process.

Balancing speed, accuracy, and completeness in research is a significant challenge. P5 discussed these competing priorities, noting the difficulty in achieving all three simultaneously. P5 explained, *“Although it’s not always possible, students can understand that being fast, accurate, and complete aren’t always compatible. So, you often need to balance these aspects.”* This insight underscores the need for researchers to make trade-offs and prioritize different aspects depending on their specific goals and constraints.

Overall, the research process is characterized by its dynamic and flexible nature, allowing for continual refinement and adjustment. Researchers must navigate the challenges of balancing speed, accuracy, and completeness, often making trade-offs to achieve their research objectives.

#### **4.6.3 Areas of Focus in the Literature Review Process**

Participants highlighted different areas of focus throughout the literature review process, reflecting the varied needs and stages of their research. This variability underscores the importance of adapting research strategies to individual expertise, needs, and preferences while maintaining a common emphasis on thoroughness and effectiveness.

The process of **categorizing and writing** is crucial for ensuring comprehensive coverage of the topic. As P1 emphasized, *“I probably focus most on categorizing and writing. Ensuring that I’m covering my topic correctly is crucial.”* This highlights the necessity of meticulous organization and synthesis of information to create a coherent and complete literature review.

The **search** process is another critical area, particularly for those supporting students in their research efforts. P2 stated, *“I primarily focus on the search because I feel it’s at the core of the librarian support for this part of the research lifecycle and related research.”* This indicates the role of mentorship in developing students’ research skills and ensuring they access high-quality information.

Guiding students towards **credible sources**, such as peer-reviewed articles, is essential, especially for inexperienced researchers. P3 noted, *“I instruct them to seek peer-reviewed articles, which can be challenging for first and second-year students who often struggle to find articles specifically addressing their topic.”* This demonstrates the role of mentorship in developing students’ research skills and ensuring they access high-quality information.

**Filtering** the most relevant papers from search results is also a key focus, as P4 explained, *“What I find most important is trying to filter out or distinguish the core papers from those that are merely related”.* This ability to discern and prioritize key literature is vital for constructing a solid foundation for any research project.

Overall, participants prioritize different aspects of the literature review process based on their roles and the specific needs of their research. Whether it is through detailed categorization, robust searching techniques, student mentorship, or discerning key literature, the common goal remains the same: to conduct thorough and effective research. By interpreting and integrating these varied focuses, a comprehensive and nuanced understanding of the literature review process emerges.

#### 4.6.4 Effectiveness of Handling the Research Processes

Participants provided insights on the effectiveness of different aspects of the research process, highlighting how students and professional researchers handle these tasks differently. Their observations emphasize the varying levels of proficiency and emphasize different parts of the research process.

**Advanced Search and Navigation Skills:** P1 and P4, both experienced researchers, emphasized the **importance of searching and navigating information**

**sources**, demonstrating their advanced skills in these areas. P1 highlighted their ability to efficiently locate relevant literature, showing the necessity of good search skills for thorough research. Similarly, P4's focus on filtering core papers from a vast pool of information highlights the critical need to identify key literature among numerous sources. Their perspectives reveal that advanced researchers prioritize and excel in effectively managing the vast amount of available information.

**Strengths in Resource Location:** P2 pointed out a common strength among students in **finding specific types of resources**, such as peer-reviewed articles. This indicates that students are often adept at using databases and search tools to find high-quality sources, which is fundamental for building a solid research foundation. P2's insights suggest that students, while proficient in locating specific resources, may still be developing the broader search strategies that seasoned researchers possess.

**Summarizing and Paraphrasing Skills:** P3 emphasized the importance of summarizing and paraphrasing, noting that some students quickly grasp these concepts. This suggests that with proper guidance, students can develop essential skills for effective literature review and academic writing, which are crucial for synthesizing information and presenting coherent arguments. P3's focus on these foundational skills indicates a need for targeted teaching methods to help students enhance their ability to process and articulate research findings.

**Time and Task Management:** P5 highlighted the variability in students' abilities, particularly in **managing time and tasks**. P5 suggested that this ability might be influenced by structured family environments, stating, *“For instance, if they had structured family lives—like being woken up early for sports or having designated times for homework—they tend to be better at scheduling and structuring their work.”* This insight points to the significant impact of early life experiences on developing effective research habits. P5's observations emphasize that personal background can play a crucial role in how well students handle the organizational aspects of research.

Overall, participants emphasized different parts of the research process based on their roles and experiences. Professional researchers like P1 and P4 focused on advanced skills such as searching and discerning key literature, highlighting the depth of experience required for effective research. Meanwhile, participants like P2 and P3 underscored the importance of foundational skills like locating peer-reviewed articles and summarizing information, which are critical for students. P5's observations on the variability in students' abilities further emphasize the need for personalized support and guidance to enhance research skills.

### 4.6.5 Challenges in Literature Review Process

Participants highlighted several challenges that both students and researchers face during the research process, revealing the complexities and difficulties inherent in conducting thorough and effective research.

P1 identified **writing the literature review** as particularly challenging, emphasizing the difficulty of organizing ideas from various sources into a coherent narrative. P1 noted, *“I find writing particularly challenging, especially the literature review”*, which highlights the conceptual demands of synthesizing information from diverse sources.

P2 highlighted two key challenges faced by students: **developing an efficient search strategy** and **critically evaluating resources**. These difficulties are often exacerbated by time constraints and the pressure to complete assignments efficiently. P2 observed, *“they often struggle with developing a good and efficient search strategy and the tools to approach it,”* indicating the need for better support and resources to help students navigate these aspects of the research process.

P3 mentioned additional challenges for students, such as **finding peer-reviewed articles** that address their specific topics and **understanding the structure of a literature review**. They emphasized the importance of helping students identify the main thesis, key discussion points, and supporting evidence. P3 explained, *“The most difficult part is often helping students understand that they don’t need to go through each sentence and make it smaller. Instead, it’s about identifying the main thesis, the key discussion points, the evidence supporting those points, and any study or methodology that supports the evidence.”* This highlights the need for guidance in structuring and synthesizing literature effectively.

P4 identified a **lack of effective tools** for extracting data and insights from research papers as a challenge for both students and researchers. Managing data can be time-consuming, and there is a need for improved tools and methods for data extraction and analysis. P4 stated, *“A challenge I observe for both students and researchers is the lack of effective tools for extracting data or insights from these papers and then building new knowledge or relating it to existing knowledge. I think the time-consuming part is managing these pieces of data.”*

P5 discussed the challenge of **balancing speed, accuracy, and completeness** in research. They emphasized that research is a skill that requires continual maintenance and practice, suggesting that students need ongoing support and guidance

to develop and refine their research skills. P5 noted, *“Although it’s not always possible, students can understand that being fast, accurate, and complete aren’t always compatible. So, you often need to balance these aspects.”* This indicates the need for students to learn how to prioritize and balance these competing demands effectively.

Overall, the participants’ insights reveal a range of challenges in the research process, from developing search strategies and evaluating resources to synthesizing information and managing data. These challenges highlight the importance of providing students and researchers with the necessary tools, support, and guidance to navigate the complexities of research successfully.

#### 4.6.6 Models and Frameworks mentioned by Participants

The participants shared diverse perspectives on their use and teaching of formal models and frameworks for conducting literature reviews.

P1 emphasizes a practical approach based on personal experience rather than strictly adhering to specific models or ontologies. However, they acknowledge the relevance of the **ACRL Framework for information literacy** [3], encouraging awareness of its existence. This suggests that while P1 prefers a more intuitive method, they recognize the value of established frameworks in guiding information literacy.

Similarly, P2 and P5 also reference the ACRL Framework, with P2 further acknowledging regional differences in frameworks for information literacy. P2 mentions the **Big6 Skills** [24] and **Bloom’s Taxonomy** [11], indicating a broad understanding of various models. They express a need to map AI literacy to classic information literacy concepts, reflecting an interest in adapting traditional frameworks to contemporary contexts. This perspective indicates the importance of evolving educational models to keep pace with technological advancements.

P3 focuses on a teaching strategy that leverages Anne Lamott’s concept of **’shitty first drafts’** [36] to encourage students to write without pressure. This approach emphasizes the importance of revision and external feedback, suggesting that the initial writing phase should be free from the constraints of perfection. P3’s method highlights the value of iterative writing and the role of feedback in the academic writing process.

In contrast, P4, despite being aware of formal methods like systematic reviews, chooses not to engage with them directly in their work. Instead, P4 emphasizes strategic searching and the concept of **undiscovered public knowledge** [59], which

involves uncovering overlooked or hidden information. They recommend dedicating 30-minute sessions to abstract reading, database testing, and material gathering without delving into full documents, saving deeper exploration for later. This approach highlights the significance of efficient research practices and strategic information gathering in developing a well-supported research project.

P5 offers a unique set of research techniques they use and teach, including the **Golden Path**, **Fishing Net approach**, and **the Bug method**. The Golden Path focuses on quickly locating authoritative sources to build a strong foundation of credible information. The Fishing Net approach encourages broad searches without focusing on specific types of materials, fostering a comprehensive understanding of the topic. The Bug method is a more advanced research technique that is especially useful for finding materials that are not easily accessible.

Overall, participants described a variety of approaches to literature review models and frameworks. While some, like P1, P2, and P5, referenced structured models such as the ACRL Framework, Big6 Skills, and Bloom's Taxonomy, others, like P3 and P4, emphasized practical techniques and strategic searching. These insights underscore the diversity in teaching and practicing literature review methodologies, reflecting a balance between formal frameworks and adaptable, real-world strategies. This diversity highlights the need for flexibility in research education, allowing for both the application of established models and the development of innovative, experience-based methods.

#### 4.6.7 Representation and Applicability of the RaMSeS Model

The RaMSeS model was generally perceived as a comprehensive representation of the literature review process. Participants acknowledged that while no model can perfectly capture every individual's approach, RaMSeS effectively mirrors their methodologies. Suggestions for enhancing the model included making it more digestible, especially for students who may initially find it complex.

P1 emphasized the utility of the RaMSeS model in capturing the essence of literature review processes and acknowledged the **thoroughness of the model**, stating, *"It's a thorough model. This model effectively captures what I do."*

P2 indicated partial alignment, recognizing familiarity with some strategies such as background knowledge, searching in databases, reading, and skimming within the model while highlighting areas they had not previously discussed, such as annotation

and using seed papers for finding more relevant articles. They stated, *“I haven’t touched on some aspects of this. But I feel that this is close to what I’ve tried to describe.”*

P3 found the model **very close to their methods** and noted the iterative nature of the research process, emphasizing the importance of background knowledge, keyword usage, and data mining from citations. P3 stated, *“This is very close to the system.”* This participant appreciated the model’s comprehensive approach, particularly in areas of search strategies and resource evaluation.

P4 echoed a positive view, noting the **comprehensiveness of the model** and its reflection on actual research processes. They mentioned, *“It’s more comprehensive than what I’ve discussed,”* and appreciated the iterative nature of the model, especially in refining keywords and filtering articles. They concluded, *“It mirrors the process quite well,”* indicating strong alignment with practical research workflows.

P5 also provided positive feedback and suggested considering multiple research styles and pathways. They stated, *“You might want to build pathways into your database search process, or at least make accommodations for all those methods to be possible”.* They appreciated features that facilitate navigation and revisiting past searches, reinforcing the model’s practicality in supporting diverse research approaches.

Overall, participants viewed the RaMSeS model as a comprehensive and flexible representation of the literature review process, effectively mirroring their methodologies. They appreciated its thoroughness and alignment with real-world research practices, though some suggested making it more digestible for students. Participants emphasized the model’s utility in capturing iterative processes and diverse research styles, reinforcing its practical applicability across various research approaches.

#### 4.6.8 The Usefulness of the RaMSeS Model on Teaching Research Methodology

The RaMSeS model was considered a valuable tool for teaching research methodologies, particularly in structuring literature reviews. Participants highlighted its potential to impart systematic approaches to students, yet emphasized the need for simplified presentation to aid comprehension, especially among beginners. Analysis of participant feedback revealed both perceived benefits and identified challenges.

P2 emphasized the potential of the model in **teaching the foundational aspects**

**of information literacy.** They stated, *“Returning to the essentials of what your model reflects, how to approach a search structurally and systematically, and how to apply critical skills, will be more important than ever.”* This highlights the model’s value in reinforcing critical evaluation and systematic search strategies, which are crucial in the context of evolving research tools and technologies.

Similarly, P4 appreciated the model’s **comprehensiveness and ease of understanding**, noting its usefulness in teaching students how to handle search results effectively. They said, *“Often, students don’t know what to do with the results of their searches.”* This suggests that the RaMSeS model can guide students in the often-overlooked steps of research, such as exploring cited references.

P5 saw the model as a **beneficial general guide**, even though they acknowledged the variability in individual research experiences. They stated, *“It could be useful and would help students recognize where they are in the process. They could assess whether they are close to the end of a particular search or if they still have more to explore.”* This indicates that the model can help students navigate the research process more systematically and recognize their progress.

Overall, the RaMSeS model was acknowledged as a valuable tool for teaching research methodologies and structuring literature reviews. Participants highlighted its potential to impart systematic approaches to students. The model was seen as beneficial for reinforcing critical evaluation and systematic search strategies (P2), helping students handle search results effectively (P4), and guiding them through the research process while recognizing their progress (P5).

#### 4.6.9 Suggested Improvements for Enhancing the RaMSeS Model as a Teaching Tool

Despite the positive feedback, some participants pointed out challenges in using the RaMSeS model, particularly concerning its complexity and applicability. They provided insightful feedback on how to improve the RaMSeS model to make it more effective as a teaching tool for the research process. Their suggestions focused on presentation, adaptability to new technologies, refinement of search techniques, and accommodating diverse research methods.

**Simplifying Presentation:** The importance of simplifying the model’s presentation was mentioned by P1 to make it more accessible, especially for students. P1 emphasized the need to turn the model into a more digestible story. They noted, *“The*

*model is good, but the presentation needs to be simplified and explained in a way that's easier to digest.*"

**Adapting Emerging Technologies:** P2 pointed out the necessity of adapting the model to incorporate emerging technologies, such as generative AI in academic databases. This indicates a future need to revise the model to stay relevant to technological advancements, ensuring it remains useful in guiding students on how to effectively utilize these new tools.

**Refining Search Techniques:** The importance of refining the 'search in database' (Fig 3.1.B) action was stressed by P4. They suggested, *"How to effectively craft searches. How to refine your search query based on the information you're gathering. That's the critical piece."* This highlights the need for the model to provide more detailed guidance on developing effective search strategies, which are crucial for comprehensive and systematic literature reviews.

P5 suggested that the model should consider different search methodologies, such as the Golden Path, the Fishing Net, and the Bug (discussed in section 4.6.6). P5 stated that *"You might want to build pathways into your database search process, or at least make accommodations for all those methods to be possible."* This shows that the model could be enriched by incorporating various search strategies within the search in databases (Fig 3.1.B), making it more versatile and applicable to a broader range of research scenarios. Additionally, they mentioned the potential benefits of features like viewing past searches which could simplify the research process and improve navigation.

**Iterative Nature of Research:** P3 highlighted the iterative nature of research, where initial findings and thesis statements may change as new information is gathered. They said, *"Your thesis may change. So you may have to go back to the beginning again to find out that a lot of the stuff you've got isn't working for your thesis as well as it could be."* This suggests that the model should emphasize the iterative process of research, allowing for flexibility and adjustments as new insights are gained.

Overall, the feedback from participants can make the model a more effective and versatile teaching tool, better-equipping students and researchers to navigate the complexities of the research process.

#### 4.6.10 Techniques and Tools for Organizing and Managing Literature Findings: Insights from Participants

The participants shared a range of techniques and tools they recommend to students for effectively organizing and managing their literature findings. Their insights highlighted the importance of using citation management software, the flexibility of different organizational methods, and the benefits of collaborative tools.

**Citation Management Software:** Participants overwhelmingly endorsed the use of citation management software, with Zotero<sup>7</sup> and Mendeley<sup>8</sup> being the most frequently mentioned tools. P1, who utilizes both Zotero and Mendeley, highlighted Zotero’s digital highlighting feature as particularly beneficial: *“In Zotero, I can automatically generate a reference from the highlighted text and possibly annotate it.”* This functionality enables users to efficiently revisit and focus on the most relevant parts of their articles.

P4 emphasized the importance of adopting a tool like Zotero early on to establish a comprehensive library of resources: *“I often recommend Zotero. This should become a routine practice, reflecting a structured approach to the work on the student’s part.”* P2 underscored the advantages of Zotero’s open-source nature and its capabilities, noting, *“I am also 100% convinced that it is one of the most capable tools available, especially for students and possibly for researchers.”* They suggested that despite initial reluctance among students to explore new software, transitioning to more sophisticated tools like Zotero can significantly enhance their research practices.

**Flexibility and Personalization:** Participants acknowledged that students have diverse preferences and processes for organizing their information. P3 encouraged students to use their own processes while incorporating additional tools: *“We all have different ways of organizing our information. I encourage students to use their own process and give them additional tools to incorporate into it.”* This flexibility is crucial as it respects individual working styles while promoting the adoption of new, more efficient methods.

P4, who uses Zotero for group projects, emphasized its collaborative features: *“Zotero’s features for group sharing and collaboration are particularly useful. We often divide the literature review into different sections and assign them to various coauthors.”*

---

<sup>7</sup><https://www.zotero.org>

<sup>8</sup><https://www.mendeley.com>

This highlights the importance of selecting tools that facilitate collaboration and streamline the literature review process in group settings.

**Alternative Tools and Methods:** Some participants mentioned alternative tools and methods that can complement citation management software. P4 noted that students conducting systematic reviews might use spreadsheets to organize variables extracted from papers. They also mentioned NVivo<sup>9</sup> as a useful tool for synthesizing qualitative data. P2 acknowledged that many students initially use simple methods like Word documents or Excel files to organize their resources.

**Physical Methods:** Despite the prevalence of digital tools, some participants still valued physical methods for certain tasks. P1 described their preference for physically underlining and writing notes in the margins of printed articles: *“I physically underline and write notes in the margins because it creates a stronger cognitive connection.”* This approach can be particularly beneficial for those who find that physical interaction with their materials enhances their understanding and retention.

Overall, participants recommended a combination of citation management software, personalized organizational methods, collaborative tools, and, in some cases, physical note-taking. Tools like Zotero and Mendeley were highly endorsed for their comprehensive features and collaborative capabilities. Flexibility in organizational methods was also emphasized, recognizing the diverse preferences and needs of students.

## 4.7 Summary and Discussion

In this section, we summarize and interpret the findings detailed in section 4.6. The discussion delves into addressing the interview research questions in Section 4.1.

The RaMSeS model is perceived by participants as a comprehensive framework that effectively represents key activities within the literature review process, which answers our first research question (RQ1, Section 4.1). Participants consistently praised its ability to navigate essential phases such as ideation, searching, organizing, and writing. They noted its alignment with their iterative research approach, crucial for refining research questions and establishing foundational perspectives (P1, P3). This initial phase was pivotal in demonstrating the model’s practical utility (P2), highlighting its role in refining and adapting research strategies continually.

---

<sup>9</sup><https://lumivero.com/products/nvivo>

Furthermore, experts lauded the RaMSeS model for its descriptive nature and its utility in developing research methodologies, addressing RQ2. They unanimously acknowledged its clarity in structuring the research process—from clarifying research questions to synthesizing findings (P1, P3). Its flexibility was particularly appreciated, allowing researchers to adapt its guidelines across diverse contexts and methodologies (P2, P4). This consensus underscores its potential contribution to enhancing systematic literature review practices, supporting researchers at all levels in their pursuit of rigorous research methodologies.

The RaMSeS model can prove to be valuable for describing processes within literature research theories and current practices, addressing RQ3. Experts highlighted its role as a comprehensive framework that people can follow. It can serve as a map for people to identify which part of the process they are in and which options they have moving forward. Participants emphasized its adaptability in accommodating dynamic shifts and methodological adaptations throughout the research process (P1, P3). This adaptability was crucial in contemporary research contexts where responsiveness to new information is important. Moreover, the model's structured approach aids researchers in navigating phases systematically, ensuring thoroughness and coherence in their literature reviews (P2, P4).

The RaMSeS model does not explicitly mention the use of Artificial Intelligence (AI). However, in our model, different actions and outputs can be accomplished by humans, machines, or a combination of both. For example, background knowledge (Figure 3.1.A) can be enriched, articles can be read/skimmed (Figure 3.1.I), and notes can be made (Figure 3.1.H) through a combination of human and machine efforts. Each action may be performed with the help of AI or entirely by AI, depending on the situation. The model remains applicable regardless of the method employed. To provide more specificity about how and who performs these actions, a closer examination and indication of roles may be necessary.

Several limitations and areas for enhancement in the RaMSeS model were noted by participants, addressing RQ4. Participants highlighted the need for more explicit guidance on integrating diverse methodologies and adapting variations in research designs (P1, P3). Suggestions were made to enhance its adaptability to emerging technologies, digital tools, and advancements in data management (P2, P4). Participants suggested that the model should support researchers in interdisciplinary fields and address challenges related to underrepresented topics (P3, P5). This includes providing guidelines for evaluating sources beyond traditional disciplinary boundaries

and promoting inclusive research practices.

## Chapter 5

# Early Career Researchers’ Management of Academic Sources

In our exploration of academic document management processes and their integration into the literature review writing, we conducted a survey study followed by an in-depth interview study, detailed in Chapter 4. The survey was designed around four pivotal research questions (Section 5.1), each aiming to unravel different facets of the academic workflow, from reading habits and note-taking practices to the maintenance of article collections and the transformation of these collections into final written reports.

In this chapter, we explain the design of the survey study, participant sample, study procedure, and Results.

### 5.1 Survey Research Questions

Research questions for the survey are crafted to comprehensively understand academic document management processes and their integration into final literature review writing. RQ1 examines reading and engagement levels pre- and post-collection. RQ2 focuses on summarization and note-taking practices. RQ3 explores collection maintenance and RMS usage. Lastly, RQ4 traces information transformation into the final report.

- **RQ1 - What are the typical researcher reading habits?** E.g., do researchers read individual articles before collecting them? How much of an article is read before and after collecting?

- **RQ2 - What are the typical summarization and note-taking practices?** E.g., when do researchers take notes or summarize?
- **RQ3 - How do researchers populate and maintain their collections?** E.g., how do researchers process the article collections? Do researchers feel that RMS supports them in processing collections?
- **RQ4 - How do researchers eventually transform what they have read and collected into writing?** E.g., to what extent do they rely on their collections when writing?

The overall goal of the survey study is to provide an initial characterization of how early career researchers accomplish the related work research process, with a scope that covers reading to writing up. To accomplish the overall goal we look at four top-level research questions.

## 5.2 Survey Design

We designed and deployed an online survey to understand the research process of early career academic researchers. Based on the main research questions (Section 5.1) and the RaMSeS model, we derived a set of survey questions (SQs) which we iteratively refined over multiple sessions. The final set of 23 questions can be divided into three groups: demographic questions that describe the research sample (SQ1-SQ7, see Table 5.2), questions to assess research practices (SQ8-SQ24, see Table 5.1), and an attention/decoy question (SQ21), which is an inverted version of SQ15. Note that some survey questions address multiple research questions and that question SQ25 is an open question that was analyzed qualitatively. The final version of the survey is estimated to take an average of 15 minutes to complete and was approved by the institutional ethics research board of our institution before deployment. The survey questions we designed and used for the study can be accessed in the Appendix A

## 5.3 Linking Survey Questions to RaMSeS and Research Questions

To explore the typical reading habits of researchers (RQ1), survey questions SQ8 and SQ9 are instrumental. SQ8 examines the time allocation across various research-

related activities, shedding light on the role of reading within the overall research process. SQ9, by asking respondents to sequence their activities following the discovery of a pivotal article, offers insight into how reading is integrated into the workflow.

For RQ2, which delves into standard summarization and note-taking practices, SQ10, SQ11, and SQ14 are pertinent. SQ10 identifies the diversity of sources from which researchers take notes, linking to their reading and summarization behaviors. SQ11 queries the methods of note-taking, directly tying to their strategies for documenting insights. SQ14 can reveal the effectiveness of these notes when researchers revisit articles, highlighting the utility of summarization for understanding and analysis.

The third research question (RQ3) examines how researchers curate and manage their personal collections of articles. SQ16 inquires about the use of archival systems and tools, providing a window into collection management strategies. SQ18 evaluates the efficiency of reference management tools, crucial for maintaining organized collections.

Investigating how researchers transform their readings and collected materials into written work (RQ4), SQ19, SQ20, and SQ22 are aligned. SQ19 aims to understand the practical application of collected information in writing. SQ20 investigates the challenges faced during this integration process. SQ22 assesses the degree to which the curated collection supports the writing endeavor.

Furthermore, SQ12 and SQ13 offer general context by assessing task prioritization and research activities, which can indirectly inform us about reading and note-taking habits. SQ15, SQ21, SQ23, and SQ24 touch upon researchers' attitudes and perceived challenges, providing a backdrop that may affect all aspects of their research practices.

Over multiple iterations, we refined our survey questions and finalized 23 questions (Table 5.1). Survey questions were designed to delve into researchers' practices, covering article search and filtering (which reflects our proposed RaMSeS model). We probed into post-article selection activities, including reading, skimming, note-taking, revisiting, analysis, and final writing. We sought insights on time allocation, evaluated the utilization of RMS tools, gauged satisfaction levels, and assessed the time invested in managing and analyzing collections. Furthermore, challenges from the initial search to the final writing stages were systematically addressed. To ensure ethical standards were upheld, we submitted our research proposal to our institution's human research ethics board, and after receiving approval, we commenced participant recruitment. The survey was approved by the local Research Ethics Board.

Table 5.1: Survey questions and associated question types.

Q #	Question Text	Question Type
SQ3	What is your gender?	Multiple Choice (4)
SQ4	What is your current age?	Number input
SQ5	What is the highest level of education you have completed or the highest degree you have received?	Multiple Choice (9)
SQ6	Which of the following best describes your current occupation?	Multiple Choice (7)
SQ7	What is your domain?	Multiple Choice (6)
SQ8	In a week, on average how much time do you spend on finding, reading, writing, manipulating, and interacting with research articles and works? (In hours)	Number input
SQ9	Suppose you have found an important article relevant to your research. Can you order the following activities in order of what you do most often?	Ranking (5)
SQ10	What type of information sources do you save in your collection?	Multiple Selection (6)
SQ11	After reading or skimming an article, do you take notes? If yes how do you take notes?	Multiple Selection (5)
SQ12	How do you rank the amount of time you spend on different research tasks?	Ranking (8)
SQ13	If you are managing a collection of references or documents, select all the possible activities you do on your collection:	Multiple Selection (8)
SQ14	During your research or literature review, how do you go back to an article?	Multiple Selection (4)
SQ15	How much do you agree with: It is easy for me to take electronic notes on various items in my collection such as PDF files, articles, etc	Range 0 (very easy) - 100 (very-difficult)
SQ16	To facilitate your research process, do you use any archiving system/tool? If yes, which of the following apply?	Multiple Selection (8)
SQ17	There are a bunch of reference management tools currently used by researchers. Do you use any of those?	Multiple Selection (13)
SQ18	If you have experience with any reference management tool, overall how well do the tools support the process of writing the final report/paper?	Range 0 (inadequate) - 100 (perfect)
SQ19	When you are writing your paper/report how do you use your collection, annotations, and notes?	Multiple Selection (6)
SQ20	In your entire process, from searching for an article to writing the final paper, which part of the process feels most challenging for you?	Ranking (7)
SQ21	How much do you agree with the following statement: It is difficult for me to take notes on various digital items in my collection such as new articles, files etc.	Range 0 (very difficult) - 100 (very-easy)
SQ22	How much do you agree with the following statement: I generally satisfactorily utilize my collection for writing the final report/paper.	Range 0 (underuse) - 100 (over-reliance)
SQ23	How much do you agree with the following statement: I wish I spent more time building my collection.	Range 0 (less time) - 100 (more time)
SQ24	How much do you agree with the following statement: I wish I spent more time analyzing my collection.	Range 0 (less time) - 100 (more time)
SQ25	Do you have any suggestions for tools, applications, or software that could improve the entire process of your research journey?	Open-ended

## 5.4 Participant Recruitment

We recruited participants through a dual approach, utilizing poster advertisements at our institution and enlisting global participants through Prolific<sup>1</sup>, an online crowd-sourced study platform. Participants received \$5 in compensation for their time. Participant screening reflects the target audience of early career researchers. Prospective participants were required to hold at least a bachelor’s degree, be currently pursuing a Ph.D., be proficient in English, and have experience in academic writing such as reports or grant applications.

Initially, we recruited 50 participants, and after applying data acceptance criteria, 32 participants qualified for further analysis. For the subsequent recruitment phase,

<sup>1</sup><https://www.prolific.com>

Table 5.2: Overall demographics of our Participants

Domain	Number	Academic Position	Age (Avg)	Research Time
Engineering/ Computer Science	25	Grad Student (23), Postdoctoral researcher (1), Professor (1)	32.24	19.72 h/w
Science	24	Grad Student (22), Postdoctoral researcher (2)	28.7	13.12 h/w
Social Science	21	Grad Student (20), Postdoctoral researcher (1)	28.14	19.38 h/w
Health Science	19	Grad Student (18), Postdoctoral researcher (1)	26.89	7.15 h/w
Arts/Humanities	8	Grad Student (7), Postdoctoral researcher (1)	30.5	19.75 h/w
Law	2	Grad Student	24	1.5 h/w
Psycholinguistics	1	Grad Student	35	2 h/w

we enlisted an additional 50 participants, resulting in 40 participants meeting the acceptance criterion. The last group of participants ( $n=28$ ) was recruited from our local institution through mailing lists and advertisements on bulletin boards. Ultimately, we gathered 100 survey valid responses. The data acceptance criteria included a minimum completion time of 8 minutes for the survey and a consistency threshold in the attention/decoy question.

The overall demographics of the analyzed sample are summarized in Table 5.2. Overall, the survey results indicate a diverse group of participants in terms of gender, (43 males, 56 females, and 1 non-binary). Age ranges from 22 to 53 years old with average age of 31.0 for males and 28.0 for females. Participants reported dedicating a considerable amount of time to research activities on a weekly basis (average weekly research time: 15.14 hours).

## 5.5 Survey Results

We report the results in this section. We describe the overall answers for SQ9-SQ24; then the results of an unsupervised cluster analysis; notable correlations between responses across survey questions and, finally; the qualitative analysis of responses to SQ25.

The use of both correlation analysis and unsupervised clustering adds depth to the analysis. Correlation analysis was chosen for its ability to reveal linear relation-

ships between variables, while unsupervised clustering was selected to uncover natural groupings within the data. Despite these advantages, correlation analysis assumes linear relationships, which may not capture all types of associations, and clustering can be sensitive to the choice of parameters. A sample size of 100 early career researchers was selected to ensure sufficient statistical power, using stratified sampling to achieve representation across various disciplines.

We found that researchers tend to prioritize skimming over research materials rather than reading in full. In addition, despite the availability of digital tools and technology, some researchers (15%) still rely on traditional physical note-taking and archiving methods for writing, knowledge refreshment, and pattern recognition. Participants (76%) using RMS expressed their satisfaction. The participants also provided suggestions for improvement and mentioned some of the features that can be implemented in the development of future tools to meet researchers' needs in terms of organization, collaboration, ease of use, integration with writing platforms, and enhanced analysis capabilities. Here we just provide a snapshot of the survey results, detailed results are discussed in the following subsections.

### 5.5.1 Overall Survey Answers

Figure 5.1 visually represents the results for the non-demographic questions. We describe the main characterization of the process grouped by the main research questions of Section 5.1. Note that, because the number of respondents ended up being exactly 100, response counts correspond also to response percentages.

#### Research Reading Habits

Responses to SQ12 show that a majority of researchers consider reading or skimming articles as well as taking notes the most time-consuming activities in the process. Reading is ranked as the first or second most time-intensive activity by 77, and taking notes and annotating articles by 36. The most common thing to do after finding a relevant article was to skim it (67 respondents ranked it as top). Reading the paper right after finding it is less common than skimming it or storing it for later reading. Although time-consuming, reading itself is not generally considered a challenging activity; only 32 respondents ranked it most or second-most challenging, and it is only the fourth most challenging on average.

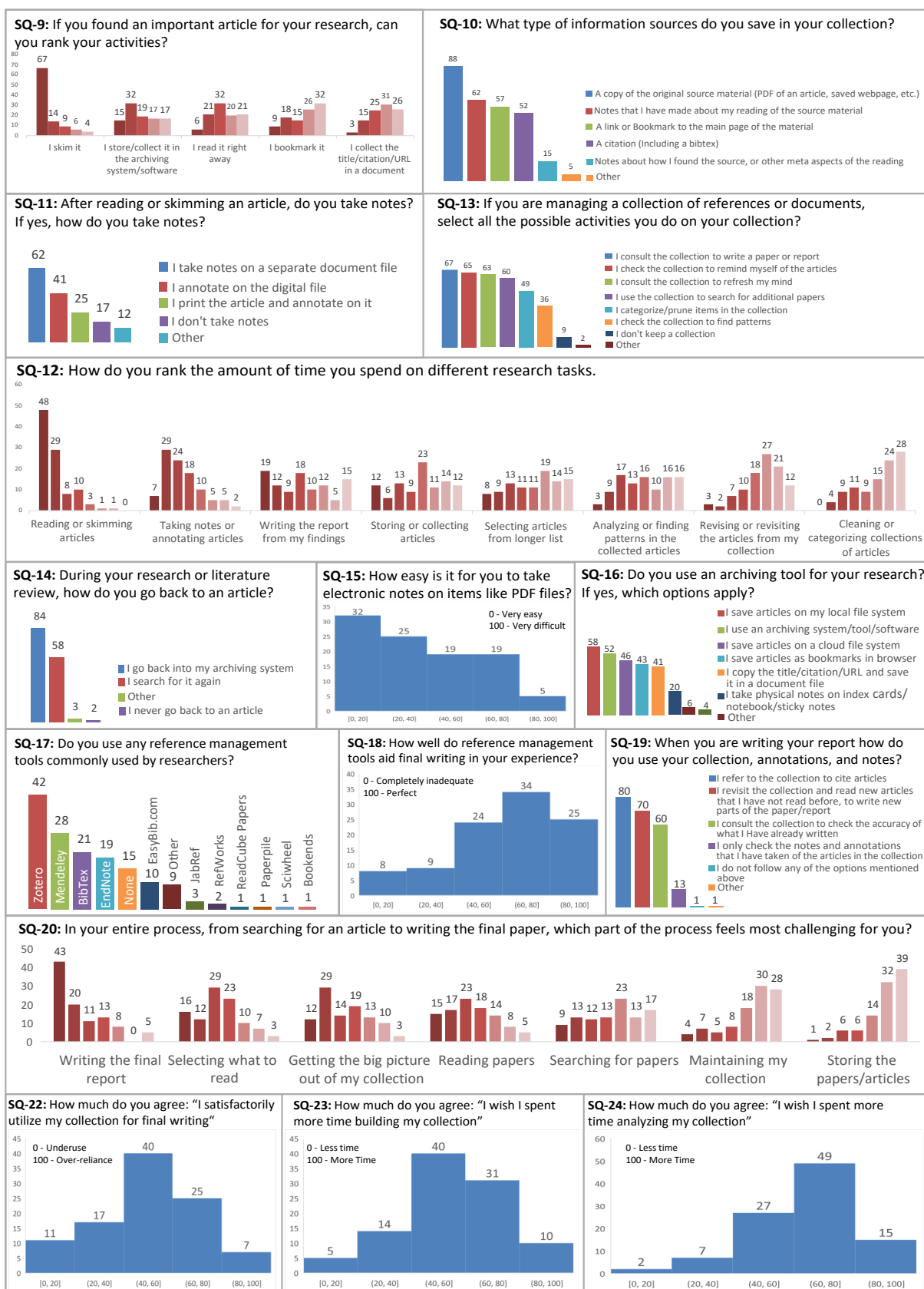


Figure 5.1: Survey Results: The answers to ranking questions (SQ9, SQ12, and SQ20) are in ascending order, with the left choices having lower cumulative averages.

## **Summarization and Note-taking Practices**

After reading or skimming articles, most people take notes (only 17 report not taking notes – SQ11). A majority report collecting notes about their reading (62 — SQ10). A majority annotate in a separate file (62), and many annotate directly on the digital file (41) or on a print (25). The relatively low proportion of people who annotate on the digital file is despite a relatively high proportion of respondents considering it easy to take electronic notes directly on the document (e.g., on the pdf file — SQ15), which is also a feature available in most reference management tools. Despite taking notes being ranked the second most time-consuming task, only 13 respondents report using it to revisit their notes (SQ19).

## **Research Work Collection Population and Maintenance**

Respondents reported a variety of materials to be collected (SQ10), including a majority who stored a copy of the article source (pdf — 88), with significant numbers of people storing also links or bookmarks, and citations. Only a few (15) store information on how the source was found. In SQ16 respondents acknowledge using a variety of storage options beyond reference management tools, which is only the second option with 52 respondents. The main alternative was saving articles directly in the local file system, which a majority of respondents report doing (58); other researchers report saving article sources on the cloud (46), as browser bookmarks (43) or as a list of citations in document files such as Word (41). Only 20 researchers use index cards or other paper-based system to keep their collection. Respondents report using a variety of reference management tools (SQ17), with the most popular being Zotero (42), followed by Mendeley (28) and EndNote (19). A substantial number (21) use BibTex, which is not a reference management tool per se, but a format to store reference information in plain text files. Storing or collecting articles was ranked fourth in terms of tasks within the process (SQ12), ranked below reading and annotating articles and writing the report; however, respondents tended to agree that they should spend more time in this part of the process (SQ23). Storing and maintaining the collection are ranked as the least challenging activities in the process (SQ20).

## **Information Transformation and Writing for Final Report**

The later stages of the process involve finding patterns in the collection and utilizing this information, as well as the collection itself, to write up the final report. Relatively

few respondents report ever checking the collected literature globally to find patterns (36 in SQ13). Respondents probably identify this as a problem, since there is, on average, a desire to spend more time analyzing the collection (SQ24 — responses are clearly above the midpoint of 50 on average). Researchers might avoid this analysis process because it is seen as challenging (“Getting the big picture out of the collection” is rated the third most challenging aspect of the process – SQ20), yet it is the sixth task in terms of how much time is spent on it (SQ12).

The writing of the report itself is the third most time-consuming task (SQ12), and the most challenging (SQ20). Respondents seem to be generally happy about the effort spent on the collection phase since most consider that they do not under-use or over-rely upon their collections for writing (SQ22). Responses to SQ19 indicate that most researchers do use the collection to cite articles when writing, to read articles that they have collected before, and to check the accuracy of statements that they make in the report. However, over 58 respondents report searching for articles again on the internet for articles, despite having them in their archiving system.

### 5.5.2 Correlation Analysis

In our study, we began with Correlation Analysis to understand basic statistical relationships between variables, identifying preliminary patterns and dependencies. This method served as a foundational step, revealing the structure and direction of associations that hinted at more complex interactions within our dataset. Recognizing the limitations of Correlation Analysis in fully capturing the intricacies of these relationships, we transitioned to Unsupervised Clustering. This advanced statistical approach allowed us to delve deeper into the data, uncovering hidden groupings and structures not evident through correlation alone. By bridging these methodologies, we moved from surface-level insights to a more nuanced understanding of our data, illustrating the evolution of our analysis from basic to complex statistical techniques.

To assess whether there were other cross-question patterns in the data we also performed a correlation analysis of responses in the different questions. We calculated Pearson Correlation Coefficients for all pairs of responses of SQ8-24 (excluding SQ15 and SQ21, which were used as attention questions) and examined those that were statistically significant. Then we looked at those with the highest correlation coefficients and discarded correlations that were structural or obvious (e.g., in questions with multiple choices the answers are naturally inversely correlated). The following



that people who take notes rely on those to recall the contents at different times. Also, people who report usage of archiving systems/tools/software (SQ16.4) also tend to report using Zotero (SQ17.1 —  $r = 0.39$ ,  $p = < 0.0001$ ). SQ16.4 also correlates positively with respondent’s satisfaction with reference management tools. SQ16.4 also correlates with a higher ranking (negative correlation because lower numeric value of ranking means higher priority) of storage/collecting behavior after finding an important article (SQ9.3  $r = -0.46$ ,  $p = < 0.00001$ ).

### 5.5.3 Unsupervised Clustering Analysis

To find patterns in how respondents answered the question (i.e., to see if there are “types of respondents” who tend to answer questions differently than others) we carried out an unsupervised clustering analysis (k-means clustering). We encoded the multiple-select and ranking question survey responses to binary values and also used the numeric responses for the survey questions (Table 5.1). This resulted in 78 feature values per participant. We set the number of clusters (k) to different values (3, 4, 5, 6, and 7) and calculated three quality metrics for each clustering result (Silhouette Coefficient [55], Davies-Bouldin index [21] and Calinski-Harabasz Index [16]). Based on these metrics we selected k=3 as it provided the best overall outcomes in those metrics (Silhouette score: 0.051, Davies-Bouldin score: 3.347, and Calinski-Harabasz score: 6.233). We performed a one-way ANOVA analysis on the clustering results to determine the most distinctive survey questions or survey question options across the three clusters. Figure 5.3 shows the top 10 most discriminative features, ranked using the F-Value (with the associated p-value being less than 0.05). The values for each group are normalized averages of the responses. We next discuss the distinctive characteristics of the three clusters, which we named **casual**, **traditional**, and **digitally-savvy** research collection managers.

#### Casual Collection Managers (Casuals)

This cluster comprises 24 respondents with an average age of 26.95 years, dedicating an average of 14.83 hours per week to research. **Casuals** are characterized by reduced note taking and a reduced likelihood of maintaining a collection to aid them in their research process. They are less likely to adopt RMSs and allocate less time to the organization of their document collections. These suggest a less structured approach to collection management, where their collections might be less organized

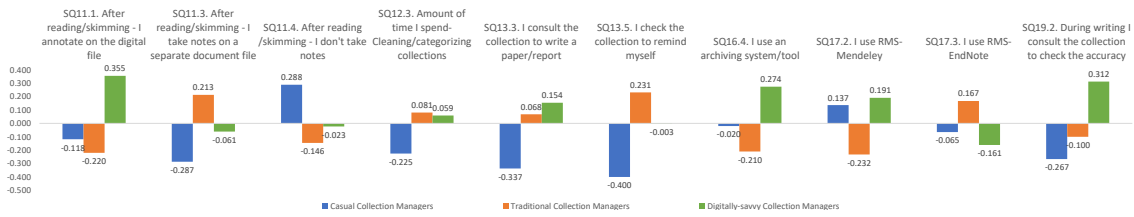


Figure 5.3: Visualization of a 3-cluster ( $k=3$ ) analysis highlighting the top 10 significant features, ranked by importance. Survey responses were encoded as binary or numeric values. The y-axis shows each cluster's average tendency to choose an option, with the x-axis set at a baseline of 0, indicating the overall average.

or non-existent, forcing them to rely more on their own memory during the writing phase. This group comprises individuals from a diverse distribution of domains, including Law (1), Science (7), Social Science (9), Health Science (5), and Engineering/Computer Science (3).

### **Traditional Collection Managers (Traditionals)**

This is the largest cluster with 42 participants with an average age of 29.57 years, dedicating an average of 14.11 hours per week to their research. Respondents in this cluster tend to take notes about the articles they read/skim and also spend time categorizing/cleaning their collection (Figure 5.3). However, they are significantly different from **Casuals** as they are more likely to take notes and specifically, notes on separate document files (Figure 5.3). In addition, this cluster of participants is also less likely to use archiving tools and systems. Among the reported RMS use, the **Traditionals** are also more less likely to use “Mendeley” and more likely to utilize “Endnote” which is an older RMS. Furthermore, they are more likely to also use their collections to remind themselves of the articles and resources they previously gathered (Figure 5.3). This group comprises individuals from a diverse array of fields, including Health Science (13), Social Science (11), Arts/Humanities (8), Engineering/Computer Science (5), Science (8), Law (1), and Psycholinguistics (1) (Health and Social Sciences are slightly over-represented compared to the other groups).

### **Digitally-Savvy Collection Managers (Digitals)**

This cluster, comprises 34 participants with an average age of 30.44 years, spending 16.61 hours per week on research. This cluster of participants is more similar to **Traditionals** than **Casuals** in that these participants have a high likelihood of maintaining and interacting with their collections. We refer to these participants as “Digitally-Savvy” due to their higher interactions with and more reported usage of digital solutions. Participants in this cluster report much higher than average likelihood to (Figure 5.3) annotate directly on digital files after reading or skimming an article. This contrasts them with the **Traditionals** cluster who also report the activity of note-taking but has a lower likelihood to annotate directly on digital files. These participants also report a higher-than-average use of archiving tools and systems (Figure 5.3). Moreover, this preference to use digital tools is matched by a higher

likelihood to consult their collections to check the accuracy when writing papers or reports than other clusters. This cluster comprises individuals from the domains of Engineering/Computer Science (18), Science (11), Arts/Humanities (2), Social Science (2), and Health Science (1) (a more dominant representation of the sciences and engineering/computer science).

#### 5.5.4 Open-ended Survey Response Analysis

SQ25 was an open question where participants provided their suggestions for tools, applications, or software that could improve the entire process of their research process. We analyzed this question qualitatively by:

- 1) extracting relevant topic mentions from the text (labeled by the participant);
- 2) grouping related mentions through affinity diagramming;
- 3) summarizing the main mentioned groups. What follows is a summary of the answers.

Participants mentioned various tools and applications they use for their research purposes. This includes Microsoft Word, Excel, EndNote, Obsidian, Zotero, Mendeley, Notion, GoodNotes, ChatGPT, The Old Reader, and Scrivener. They highlight different purposes for using these tools, such as organizing documents, notes, and references, taking digital notes directly on articles, storing and annotating papers, creating links between ideas and sources, sorting and cross-referencing files, and building a holistic overview of their research collection. One participant (P82) expressed a desire for tools that could further support collaborative research teams. P72 wanted the ability to extract text based on categories and enable the analysis and collaboration of RMS.

While some participants expressed satisfaction with their current tools, others mentioned specific limitations. These included the need to manually format changes in EndNote, the difficulty of managing a large number of references in existing tools, slow performance, and frequent crashes. Participants also identified challenges in taking online article notes, retrieving all notes from a single article tag in Zotero, and establishing a coherent and consistent note-taking and collection system. Additionally, they expressed a desire for tools that would update and maintain collections after projects were completed.

Respondents provided suggestions for improving existing tools and developing new ones. These suggestions include the integration of research software with Word

for accessibility on any device (P96), the ability to search and save articles from Google Scholar into a personal library (P11), and easier digital note-taking directly on articles (P14). Other suggestions include a single sheet for accessible studies with visible graphs and links (P16), improved integration with iPads for handwritten notes (P20), and hover-over summaries for article titles. They also suggested more visually appealing and interactive note-taking interfaces (P89), software that matches similar notes from different papers, and simplified implementation with Word software.

Some responses focused on enhancing RMS by allowing customizable inputs for notes sections (P35), providing complete journals in database listings, creating tools resembling outlines for a better understanding of folders and notes, and incorporating AI-based features for sorting papers by results and collecting/archiving reading material (P43). Participants also emphasized the importance of software that draws common themes from multiple papers, displays graphical relationships between papers or authors (P80), and supports collaborative reference management for research teams (P72).

## 5.6 Summary and Discussion

Here, we summarize and interpret the findings detailed in the results in section 5.5. The responses offer a general characterization of early career researchers' practices regarding the research questions of Section 5.1 and the actions and outputs delineated in the model described in Section 3. We refer to elements of the model based on the nomenclature from Figure 3.1, e.g., M.G refers to the "collect" action.

The survey provides evidence that most early career researchers apply effective reading techniques when they find relevant papers, preferring to first skim rather than read articles fully (RQ1). This is likely to make it even more important for the diligent collection of the reference or the article itself, i.e., the steps between M.F-M.G-M.J, so that the paper can be effectively retrieved when needed again for a full read or for citing. Most researchers report taking notes (RQ2), and the ones that do seem to have a system for organizing their notes. However, there appears to be a significant overlap between researchers who do not take notes and those who do not systematically manage their collections, a group referred to as the 'casuals' (RQ3). Interestingly, while taking notes is common, the use of those notes in practice is less frequent.

Interestingly, taking notes is much more common than using those same notes to

remind oneself of what has been read (RQ2). This might be a missed opportunity, or simply that the researchers recognize the value of the act of taking notes in itself, separately from the value of accessing those notes afterward.

Researchers adopt diverse strategies to keep the source material in their collection (consistent with [65] — RQ3). These techniques include copying the sources, taking notes, and managing links and citations. The largely overlapping responses suggest that many respondents use multiple systems or are in transition.

Participants reported a high affinity for saving copies of the original material (SQ10) (38 out of 42 **Traditionals**, 19 out of 24 **Casuals**, and 31 out of 34 **Digitals**). According to SQ10, a large majority, 88/100 participants (38 out of 42 (**Traditionals**), 19 out of 24 (**Casuals**), and 31 out of 34 (**Digitals**)) opted to save copies of the original source material, reflecting a very common practice. 27 out of 42 **Traditionals**, 8 out of 24 **Casuals** and 27 out of 34 **Digitals** store personal notes after reading the source material. 27 out of 42 **Traditionals**, 8 out of 24 **Casuals**, and 27 out of 34 (**Digitals**) indicate a preference for saving links or bookmarks. The **Casuals** participants consistently report lower note-taking and research material-saving behaviors. This varied range of practices highlights the complex nature of research material curation and implies that researchers use a **combination of techniques** to sort and organize their article collections.

Respondents seem to recognize the value of existing reference management tools as facilitators to create collections, and their use seems to be considered easy (RQ3). An exception is the **Casuals**, who tend to favor saving links or bookmarks (probably the most brittle way to keep references). Simultaneously, overall, respondents do not generally pay much heed to the maintenance of their collections (RQ3). In terms of the model, this suggests that existing tools or processes are fairly effective in the collection process (M.F-M.G-M.J.). A minor additional point is that most respondents do not seem to record meta-search information (e.g., how they found a particular paper). Because this is information that most tools do not collect either, there might be an opportunity here for new features that support provenance and more systematic exploration of literature (e.g., recording keywords used to find a paper or connecting papers that have been found on the same day or through the same search).

Perhaps more interestingly, the processes of revisiting articles and notes, cleaning and categorizing the collection utilizing the reference collection effectively, and generally extracting the sense of the collection itself (M.K and M.L, RQ4) seem to be much more poorly supported by existing tools. These processes are still considered

difficult. Respondents also reported spending little time on those, while generally recognizing that spending more time working with the collection might be beneficial. Additionally, we find it interesting that, despite generally collecting references and having systems to store articles, more researchers than we had predicted do not use those collections for much in the later processes. Many even report searching again on the internet to access materials that they have seen again. We speculate that this is because current RMSs and other ad hoc systems that researchers use are likely worst at search, especially if the collection is large.

This points to opportunities for improvement of existing tools. For example, RMSs could have better searches, better ways to highlight or “star” papers, or effectively communicate which papers are more important because they have been used more, re-read, etc. However, there are also opportunities for new tools that enhance the more difficult and cognitively demanding parts of the process: tasks involving sense-making and pattern recognition of large collections as well as translating the newly acquired knowledge about research in an area into the final text (i.e., a background section, related work section, survey paper, etc.). This is also supported by the comments and feature requests reported by respondents in SQ25.

We have seen a recent explosion of tools taking advantage of new machine learning technologies, including large language models, for finding literature (e.g., ResearchRabbit<sup>2</sup>), but we believe that the actions in the lower third of the model have been generally neglected despite their importance, and ML-based techniques are not the only way to address this need. For example, we find the prospect of tools that provide enhanced visualization of reference networks, more intuitive methods for selecting relevant literature, and streamlined organization of citations during writing to be particularly promising. These tools could greatly assist researchers in navigating complex information landscapes, making connections between diverse sources, and ultimately producing more coherent and well-supported academic work.

Our clustering analysis also offers a glimpse of the diversity of current approaches to the research process. The existence of a significant group of researchers (the [Casuals](#)), who do relatively little storage and tracking of sources, may be due to various factors. These could include a perception that current tools offer limited benefits (with some relying on memory and re-retrieval from the internet as sufficient strategies) or workflows that are incompatible with existing tools. Further investigation into why this group may not engage in more systematic approaches to source management

---

<sup>2</sup><https://www.researchrabbit.ai/>

is necessary and important.

The other two groups are more search-literate but have their own idiosyncrasies. The **Traditionals** are more likely to use long-tested tools such as EndNote, but also to collect notes separate from the documents themselves. They also seem to be the cluster most invested in using their collection afterward. The **Digitals** take advantage of more current tooling, but seem, on average, to leverage their collection less. It's important to note that these observations primarily apply to research processes where RMS can be effectively utilized. For other types of research, such as archival or field research, RMS may offer limited assistance, reflecting the diverse research practices present in research-centric institutions.

The dominance in our sample of the **Traditionals**, who also seems to be less connected to health sciences and social sciences and humanities than the **Digitals** (more skewed towards the computer science and engineering disciplines) might reflect the slower pace of penetration of new tools in those disciplines. Nevertheless, the **Traditionals** seem to get better value from their collections. It is possible that researchers who rely more on fully digital workflows also miss out on parts of the process that are less well supported by current software (the bottom third of the RaMSeS model of Figure 3.1).

## Chapter 6

# Discussion and Future Directions

In this chapter, we interpret the findings from the RaMSeS model validation interview study (Chapter 4) and early career researchers' management of academic sources survey study (Chapter 5). We explore the alignment and contradictions between the two studies, the implications of the findings for both novice and experienced researchers, and the potential developments in software that could enhance the research process. We also identify future research directions and discuss the limitations of our current knowledge.

*The RaMSeS model validation interview study* revealed that participants found the RaMSeS model to be a comprehensive framework for the literature review process, aligning well with iterative research approaches. The model was appreciated for its flexibility and clarity in structuring research methodologies. Participants noted its value in navigating phases systematically and highlighted a need for more explicit guidance on integrating diverse methodologies and adapting to new technological advancements.

*The early career researchers' management of academic sources survey study* indicated that most early career researchers employ effective reading techniques, initially skimming papers and diligently collecting references for later use. Note-taking was common, although many researchers did not frequently revisit their notes. There was a variety of strategies for curating research materials, including saving copies of original sources, taking notes, and managing links and citations. Existing reference management tools were generally well-regarded for collection purposes but were less effective for processes like revisiting, categorizing, and utilizing references.

*Alignments and Discrepancies Between Studies:* Both studies highlighted the iterative nature of the research process and the importance of systematically organizing

and revisiting research materials. The value of structured methodologies and flexible tools was emphasized in both studies, with participants recognizing the benefits of models and frameworks like RaMSeS for guiding the research process. The survey and interview findings both pointed to the need for improved tools that support the entire research process, from initial ideation to final write-up.

The RaMSeS model appears to encapsulate the entire research journey effectively, with participants commending its comprehensiveness in the interview study. However, the survey study indicates that existing tools are not adequately equipped to support the later stages of research, such as revisiting and utilizing references. The survey study also suggested that many researchers do not maintain their collections systematically, a point less emphasized in the interviews, where participants seemed more focused on the strategic and methodological aspects of the research process.

*Implications for Researchers:* For early career researchers engaged in traditional academic work, the findings suggest that structured frameworks like RaMSeS could be beneficial, providing clear guidelines for organizing and conducting literature reviews. While many early-career researchers already demonstrate strong capabilities in these areas, adopting systematic note-taking and reference management habits early in their careers could further enhance the efficiency and effectiveness of their research processes.

For experienced researchers, the findings indicate that refining existing methodologies by incorporating elements of the RaMSeS model, particularly its emphasis on iterative refinement and strategic adaptation, could be valuable. The need for tools that better support advanced stages of research, such as synthesizing findings and writing, is critical for experienced researchers who manage larger and more complex collections of research materials.

*Recommendations for Software Development:* The findings suggest several implications for software development. Extensions to existing software could include better search capabilities, tools for highlighting and categorizing important papers, and systems for tracking the use and re-use of references. New software solutions could focus on the more cognitively demanding aspects of research, such as sense-making, pattern recognition, and translating research findings into coherent written outputs. Integrating AI capabilities to assist with reading, note-taking, and synthesis could significantly enhance the efficiency and effectiveness of the research process.

*Limitations:* The survey study primarily focuses on early-career researchers and might not fully capture the experiences of more seasoned researchers or those from

specific disciplines. Additionally, the interview sample was limited to a specific geographic area and consisted of a small number of participants. To gain a more comprehensive understanding, it would be beneficial to include researchers from other countries and a larger participant pool in future studies.

## 6.1 Future Directions

Our work extends beyond the immediate findings; it sets the stage for future research and innovation in the realm of academic document collection and curation. The insights and classifications derived from our study provide a foundational understanding of the diverse methods employed by early career researchers. This understanding not only informs the development of next-generation tools and methodologies but also highlights the dynamic and evolving nature of research processes. For example, our discovery that researchers under-utilize their notes can inform the design of future tools with enhanced note-taking functionalities. These advancements aim to better support researchers in managing and utilizing their academic sources effectively, thereby improving overall research productivity and scholarly output.

Future research could delve deeper into the reasons why casual researchers may be hesitant to adopt more systematic approaches. Understanding these barriers can guide the design of training programs that effectively address and overcome these challenges. Additionally, exploring the integration of emerging technologies, such as Generative AI and machine learning, at various stages of the research process could yield valuable insights. Developing visualization tools to help researchers manage and interpret large collections of references and notes could significantly benefit the research community. Investigating how the RaMSeS model and similar frameworks can be adapted to support interdisciplinary research could promote inclusive practices and address underrepresented topics.

# Chapter 7

## Conclusion

We provided insights into the research process of early career researchers. Our approach started with the introduction of a model in Chapter 3 that delineates the academic document and reference collection management process within the context of producing academic works. In Chapter 4, we discussed the findings from the semi-structured interview study conducted to validate our proposed model. We utilized the model to design and deploy an online survey to gain insights into how early career researchers manage their large-scale documents and whether the current tools available adequately address their needs. With this wealth of information, we provided important results and classifications of different researchers concerning their research processes in Chapter 5.

The survey study was conducted before the RaMSeS model validation interview study. The survey results and findings motivated us to delve deeper into the validation of the model, leading us to the semi-structured interview study. The RaMSeS validation study strengthens our understanding of the literature management and curation process, making the model a teaching tool and a reference for developers in their current and future work. The RaMSeS model's flexibility and comprehensiveness make it a valuable reference for both educators and developers in their ongoing efforts to enhance research methodologies.

The RaMSeS model is seen as a valuable tool for teaching research methodology. It covers most of the research literature phases employed by researchers. Writing was identified as a challenging yet crucial phase, involving the synthesis of collected information into a coherent narrative. Participants viewed the RaMSeS model as a comprehensive and flexible representation of the literature review process, effectively mirroring their methodologies.

Our survey study characterizes early career researchers' research processes with respect to academic sources and citations. We propose the RaMSeS model of activities and outcomes and report the analysis of an online survey to collect insights from early career researchers managing collections of academic sources. The results show a varied landscape of processes, which seem to be better supported in the earlier stages (search, collecting references) than in the harder later stages (making sense of the collections, revisiting, and converting the knowledge into text). The cluster analysis also illuminates patterns of use by three main groups: Casual, Traditional, and Digitally-savvy collection managers. The survey results also show that researchers tend to prioritize skimming over research materials rather than reading in full. Additionally, despite the availability of digital tools and technology, some researchers (15%) still rely on traditional physical note-taking and archiving methods for writing, knowledge refreshment, and pattern recognition. Participants (76%) using RMS expressed their satisfaction.

This work contributes to understanding the evolving nature of research processes and helps design the next generation of tools. Our efforts have yielded insights, and significant results, and classified researchers into distinct categories of early career academic researchers: Casual Collection Managers, Traditional Collection Managers, and Digitally-Savvy Collection Managers. This classification enriches our understanding of research processes, offering nuanced perspectives for effective academic navigation. Furthermore, our findings and discussions have charted new directions and challenges for academic document collection creation and curation.

# Bibliography

- [1] Mortimer J Adler and Charles Van Doren. *How to read a book*. Simon and Schuster, 1972.
- [2] Sönke Ahrens. *How to take smart notes: One simple technique to boost writing, learning and thinking*. Sönke Ahrens, 2022.
- [3] American Library Association. Framework for information literacy for higher education, feb 2015. Accessed: 2024-06-02.
- [4] P. Armstrong. Bloom’s taxonomy, 2010. Retrieved August 16, 2024.
- [5] Kumaripaba Athukorala, Eve Hoggan, Anu Lehtiö, Tuukka Ruotsalo, and Giulio Jacucci. Information-seeking behaviors of computer scientists: Challenges for electronic literature search tools. *Proceedings of the American Society for Information Science and Technology*, 50(1):1–11, 2013.
- [6] Muhammad Basri and Andi Anto Patak. Exploring indonesian students’ perception on mendeley reference management software in academic writing. In *2015 2nd international conference on information technology, computer, and electrical engineering (ICITACEE)*, pages 8–13. IEEE, 2015.
- [7] Charles Bazerman. What do sociocultural studies of writing tell us about learning to write. 2016.
- [8] Nicholas J Belkin and Alina Vickery. Interaction in information systems: A review of research from document retrieval to knowledge-based systems. (*No Title*), 1985.
- [9] David C Blair et al. Information retrieval. *Journal of the American Society for Information Science*, 30(6):374–375, 1979.

- [10] Catherine Blake and Wanda Pratt. Collaborative information synthesis i: A model of information behaviors of scientists in medicine and public health. *Journal of the American Society for Information Science and Technology*, 57(13):1740–1749, 2006.
- [11] Benjamin Samuel Bloom, Max D Engelhart, Edward J Furst, Walker H Hill, and David R Krathwohl. *Taxonomy of educational objectives*, volume 2. Longmans, Green New York, 1964.
- [12] Wayne C Booth, Gregory G Colomb, and Joseph M Williams. *The craft of research*. University of Chicago press, 2009.
- [13] George EP Box. Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799, 1976.
- [14] Andrei Broder. A taxonomy of web search. In *ACM Sigir forum*, volume 36, pages 3–10. ACM New York, NY, USA, 2002.
- [15] Tongan Cai, Chacha Chen, Ting-Hao Huang, and Frank E Ritter. What makes a good reference manager? a quantitative analysis of bibliography management applications. In *Asian CHI Symposium 2021*, pages 64–69, 2021.
- [16] Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.
- [17] Donald Owen Case. Collection and organization of written information by social scientists and humanists: a review and exploratory study. *Journal of information science*, 12(3):97–104, 1986.
- [18] Catherine Chavula, Yujin Choi, and Soo Young Rieh. Searchidea: An idea generation tool to support creativity in academic search. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*, pages 161–171, 2023.
- [19] Victoria Cole and Mish Boutet. Researchrabbit. *The Journal of the Canadian Health Libraries Association*, 44(2):43, 2023.
- [20] Julie Courraud. Zotero: A free and open-source reference manager. *Medical Writing*, 23(1):46–48, 2014.

- [21] D Davies and D Bouldin. A cluster separation measure: Ieee transactions on pattern analysis and machine intelligence. itpidj 0162-8828, pami-1, 2 224–227. *Crossref Web of Science*, 1979.
- [22] Jamie DeCoster. Meta-analysis notes. 2004.
- [23] Michael B Eisenberg. Information literacy: Essential skills for the information age. *DESIDOC journal of library & information technology*, 28(2), 2008.
- [24] Michael B Eisenberg and Robert E Berkowitz. *Information Problem Solving: The Big Six Skills Approach To Library & Information Skills Instruction*. ERIC, 1990.
- [25] David Ellis. A behavioural approach to information retrieval system design. *Journal of documentation*, 45(3):171–212, 1989.
- [26] Eugene Ashley High School. The big 6 - research process - libguides at eugene ashley high school, n.d. Accessed: August 16, 2024.
- [27] Gene V Glass. Primary, secondary, and meta-analysis of research. *Educational researcher*, 5(10):3–8, 1976.
- [28] Bradley M Hemminger, Dihui Lu, KTL Vaughan, and Stephanie J Adams. Information seeking behavior of academic scientists. *Journal of the American society for information science and technology*, 58(14):2205–2225, 2007.
- [29] Christoph Hölscher and Gerhard Strube. Web search behavior of internet experts and newbies. *Computer networks*, 33(1-6):337–346, 2000.
- [30] Sharon Ince, Christopher Hoadley, and Paul A Kirschner. The role of libraries in teaching doctoral students to become information-literate researchers: a review of existing practices and recommendations for the future. *Information and Learning Sciences*, 120(3/4):158–172, 2019.
- [31] Alireza Isfandyari-Moghaddam and Vahideh Kashi-Nahanji. Does information technology affect the level of information literacy? a comparative case study of high school students. In *Aslib Proceedings*, volume 63, pages 618–631. Emerald Group Publishing Limited, 2011.

- [32] Heidi Julien and Shelagh K Genuis. Librarians' experiences of the teaching role: A national survey of librarians. *Library & Information Science Research*, 33(2):103–111, 2011.
- [33] Arunava Kali. Reference management: A critical element of scientific writing. *Journal of Advanced Pharmaceutical Technology & Research*, 7(1):27, 2016.
- [34] Ian A Knight, Max L Wilson, David F Brailsford, and Natasa Milic-Frayling. Enslaved to the trapped data: A cognitive work analysis of medical systematic reviews. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, pages 203–212, 2019.
- [35] Carol C Kuhlthau. Inside the search process: Information seeking from the user's perspective. *Journal of the American society for information science*, 42(5):361–371, 1991.
- [36] Anne Lamott. Shitty first drafts. *Writing about writing: A college reader*, pages 527–31, 1994.
- [37] Elaine Leong. Read. do. observe. take note! *Centaurus*, 60(1-2):87–103, 2018.
- [38] Ying-Hsang Liu, Paul Thomas, Tom Gedeon, and Nicolay Rusnachenko. Search interfaces for biomedical searching: How do gaze, user perception, search behaviour and search performance relate? In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*, pages 78–89, 2022.
- [39] Annemaree Lloyd. Information literacy: different contexts, different concepts, different truths? *Journal of Librarianship and information Science*, 37(2):82–88, 2005.
- [40] Stephann Makri and Claire Warwick. Information for inspiration: Understanding architects' information seeking and use behaviors to inform design. *Journal of the American Society for Information Science and Technology*, 61(9):1745–1770, 2010.
- [41] Gary Marchionini. *Information seeking in electronic environments*. Number 9. Cambridge university press, 1995.
- [42] Gary Marchionini. Search, sense making and learning: closing gaps. *Information and Learning Sciences*, 120(1/2):74–86, 2019.

- [43] David Matthews. Drowning in the literature? these smart software tools can help. *Nature*, 597(7874):141–142, 2021.
- [44] L Mercadé, FJ Díaz-Fernández, M Sinusia Lozano, J Navarro-Arenas, M Gómez, E Pinilla-Cienfuegos, D Ortiz de Zárate, VJ Gómez Hernández, and A Díaz-Rubio. Research mapping in the teaching environment: Tools based on network visualizations for a dynamic literature review. In *INTED2023 Proceedings*, pages 3916–3922. IATED, 2023.
- [45] Victor Minichiello. In-depth interviewing: Researching people. (*No Title*), 1990.
- [46] Mario Alberto Moreno Rocha, Miguel A Nacenta, and Juan Ye. An exploratory study on academic reading contexts, technology, and strategies. In *Proceedings of the 8th Mexican Conference on Human-Computer Interaction*, pages 1–10, 2021.
- [47] Paul Muter and Paula Maurutto. Reading and skimming from computer screens and books: The paperless office revisited? *Behaviour & information technology*, 10(4):257–266, 1991.
- [48] Sheshera Mysore, Mahmood Jasim, Haoru Song, Sarah Akbar, Andre Kenneth Chase Randall, and Narges Mahyar. How data scientists review the scholarly literature. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*, pages 137–152, 2023.
- [49] Ilias Nitsos, Afrodite Malliari, and Rodopi Chamouroudi. Use of reference management software among postgraduate students in greece. *Journal of Librarianship and Information Science*, 54(1):95–107, 2022.
- [50] Xi Niu, Bradley M Hemminger, Cory Lown, Stephanie Adams, Cecelia Brown, Allison Level, Merinda McLure, Audrey Powers, Michele R Tennant, and Tara Cataldo. National study of information seeking behavior of academic researchers in the united states. *Journal of the American Society for Information Science and Technology*, 61(5):869–890, 2010.
- [51] Tor Ole B Odden and Rosemary S Russ. Defining sensemaking: Bringing clarity to a fragmented theoretical construct. *Science Education*, 103(1):187–205, 2019.
- [52] Kyong Eun Oh. *The process of organizing personal information*. Rutgers The State University of New Jersey, School of Graduate Studies, 2013.

- [53] Thomas B Passin. Browser bookmark management with topic maps. In *Extreme Markup Languages*(<sup>®</sup>). Citeseer, 2003.
- [54] Antje Proske, Christina Wenzel, and Manuela Barbara Queitsch. Reference management systems. In *Digital Writing Technologies in Higher Education: Theory, Research, and Practice*, pages 215–230. Springer, 2023.
- [55] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.
- [56] Nirmal Roy, Felipe Moraes, and Claudia Hauff. Exploring users’ learning gains within search sessions. In *Proceedings of the 2020 conference on human information interaction and retrieval*, pages 432–436, 2020.
- [57] Rutgers School of Communication and Information. Information search process by carol kuhlthau. Retrieved August 16, 2024.
- [58] Marie Speare. Graduate student use and non-use of reference and pdf management software: an exploratory study. *The Journal of Academic Librarianship*, 44(6):762–774, 2018.
- [59] Don R Swanson. Undiscovered public knowledge. *The Library Quarterly*, 56(2):103–118, 1986.
- [60] Jaime Teevan, William Jones, and Robert Capra. Personal information management (pim) 2008. In *ACM SIGIR Forum*, volume 42, pages 96–103. ACM New York, NY, USA, 2008.
- [61] Gareth Terry, Nikki Hayfield, Victoria Clarke, Virginia Braun, et al. Thematic analysis. *The SAGE handbook of qualitative research in psychology*, 2(17-37):25, 2017.
- [62] Renata Tesch. *Qualitative research: Analysis types and software*. Routledge, 2013.
- [63] David A Thompson. Interface design for an interactive information retrieval system: A literature survey and a research system description. *Journal of the American Society for Information Science*, 22(6):361–373, 1971.

- [64] Hapnes Toba, Christopher Starry Jomei, Lotanto Setiawan, Oscar Karnalim, and Hui Li. Predicting users' revisitatioin behaviour based on web access contextual clusters. In *2020 8th International Conference on Information and Communication Technology (ICoICT)*, pages 1–6. IEEE, 2020.
- [65] Jesús Tramullas, Ana I Sánchez-Casabón, and Piedad Garrido-Picazo. Studies and analysis of reference management software: A literature review. *arXiv preprint arXiv:1508.07480*, 2015.
- [66] Raymond Uzwysyn. Academic libraries and technology: an environmental scan towards future possibility. 2018.
- [67] Pertti Vakkari. Searching as learning: A systematization based on literature. *Journal of Information Science*, 42(1):7–18, 2016.
- [68] Teena Willoughby, S Alexandria Anderson, Eileen Wood, Julie Mueller, and Craig Ross. Fast searching for information on the internet to use in a learning context: The impact of domain knowledge. *Computers & Education*, 52(3):640–648, 2009.
- [69] Tom D Wilson. Models in information behaviour research. *Journal of documentation*, 55(3):249–270, 1999.
- [70] Rachel Rains Winslow, Sarah L Skripsky, and Savannah L Kelly. Not just for citations: Assessing zotero while reassessing research. *Information literacy: Research and collaboration across disciplines*, pages 287–304, 2016.
- [71] Shazia Yasmeeen, Irfan Ali, and Nosheen Fatima Warraich. Personal digital information management practices of engineering faculty: finding, organizing, and re-finding information. *Pakistan Journal of Information Management & Libraries*, 21:88–103, 2019.
- [72] Yan Zhang. College students' uses and perceptions of social networking sites for health and wellness information. *Information Research: An International Electronic Journal*, 17(3):n3, 2012.
- [73] Yingting Zhang. Comparison of select reference management tools. *Medical reference services quarterly*, 31(1):45–60, 2012.

# Appendix A

## Additional Information

We include the Survey Questions from the Survey Study, Script and Codebook for the Interview Study and the Post-interview Demographics Survey questions.

**Survey Study Questions:** The survey questions including decoys.

**Interview Study Script:** The complete script of the evaluation. Note that the script in orange was action items for the experimenter.

**Interview Study Codebook:** The codebook was compiled by the primary researcher only. The review of the codebook was conducted by a three-member research team.

**Post-interview Demographics Survey:** To collect demographics of Interview participants for overall results.

# Survey Study Questions



## Meta-search Information Practices and Support Participant Consent Form

You are invited to participate in a study titled "Meta-search Information Practices and Support" that is being conducted by [REDACTED]. He is a master's student in the Computer Science department at the University of Victoria. You may contact him if you have further questions at [REDACTED]

As a master's student, he is required to conduct research to complete his MSc degree in Computer Science. The study is being developed under the supervision of Dr. Miguel Nacenta. You may contact his supervisor, [REDACTED] at [REDACTED]. The study is funded by a grant from Natural Sciences and Engineering Research Council (NSERC).

**Purpose and Objectives**

The purpose of this research project is to understand how students or researchers use tools/software or apps to facilitate their final report writing.

**Importance of this Research**

The knowledge that we gather through this research can help us and others recommend specific ways and features for the tools/software that can be presented that benefit student, academics, and researchers. Additionally, they might help us, and others design tools that will be beneficial for their use.

**Participants Selection**

You are being asked to participate in this study because you are a university student, academic, or researcher.

**Power Relationship**

The lead experimenter Mohammad Shakirul Islam will be the only one knowing about your participation, and able to access the participation data under normal circumstances. Miguel Nacenta will not participate in recruitment or send out emails to students in their course if they are the course instructor. Miguel Nacenta will not know who has enrolled in the study.

**What is involved**

If you consent to voluntarily participate in this research, your participation will include

- Reading this consent declaration
- Completing a Survey that contains a few questions related to you and the study

We estimate that the survey will take 10-15 minutes of your time.

**Inconvenience**

None beyond the time spent needed to participate in the study.

**Risks**

There are no known or anticipated risks to you by participating in this research.

**Benefits**

The potential benefits of your participation in this research include benefits to the state of knowledge, particularly in human-computer interaction.

**Compensation**

To compensate you for your participation for your time, we will provide a CAD\$5 gift card or transfer to your bank with a suitable payment medium. You will receive the amount after completing the survey. We encourage you to complete the survey properly.

**Voluntary Participation**

Your participation in this research must be completely voluntary. If you do decide to participate, you may withdraw at any time without any consequences or explanation. If you do withdraw from the study, it is logistically impossible to remove individual participant data once the survey is submitted.

### **On-going Consent**

This letter of consent only applies to this survey. You are allowed to withdraw from the study at any time without any negative consequences. No identifiable information about you will be collected.

### **Anonymity**

In terms of protecting your anonymity, the survey is completely anonymous. Participants will be anonymous in the dissemination of results. You can also receive the informed consent form via email. We will also ask you to provide your email at the end of the survey if you are interested in participating in any follow-up studies. This is completely voluntary.

### **Confidentiality**

Your confidentiality and the confidentiality of the data will be protected by access restricted, and password-protected computers located in a research lab at the University of Victoria. Only the research investigators will be able to access the data.

### **Dissemination of Results**

It is anticipated that the results of this study will be shared with others in the following ways: thesis, dissertation, class presentations, published articles, chapters or books, presentations at scholarly meetings, internet.

### **Commercial Use of Results**

There is no commercial use of data collected during this study.

### **Disposal of Data**

Data from this study will be destroyed 5 years after the beginning of the study. All data collected in the study and backups will be electronically deleted; physical documents will be shredded.

### **Contacts**

Individuals that may be contacted regarding this study include [REDACTED], [REDACTED], and [REDACTED]. In addition, you may verify the ethical approval of this study, or raise any concerns you might have, by contacting the [REDACTED].

By completing and submitting the survey, YOUR FREE AND INFORMED CONSENT IS IMPLIED and indicates that you understand the above conditions of participation in this study and that you have had the opportunity to have your questions answered by the researchers.

You may **download** the Participant Consent Declaration in [PDE](#).

\* 1. Do you understand the conditions of participation in this study and agree to participate?

Yes

No

## Meta-search Information Practices and Support Participant ID

Please input your Participant ID, received from the research team.

2. Please input your Participant ID?



## Meta-search Information Practices and Support Demographic Questions

\* 3. What is your gender?

Female

Male

Prefer not to say

Other (please specify)

\* 4. What is your current age?

\* 5. What is the highest level of education you have completed or the highest degree you have received?

- Associate degree or diploma
- Undergraduate student
- Bachelor degree
- Master's student
- Master's level graduate degree
- Doctoral student
- Doctorate level graduate degree
- Post-graduate studies or appointments
- Other (please specify)

\* 6. Which of the following best describes your current occupation?

- Undergraduate Student
- Grad Student
- Professor
- Researcher
- Postdoctoral researcher
- Librarian

Other (please specify)

\* 7. What is your domain?

## Meta-search Information Practices and Support

### Questions related to the study

\* 8. In a week, on average how much time do you spend on finding, reading, writing, manipulating, and interacting with research articles and works? (In hours)

\* 9. Suppose you have found an important article relevant to your research. Can you order the following activities in order of what you do most often? (1- means most often) [You may drag and drop the options]

☰	▾	I read it right away
☰	▾	I skim it
☰	▾	I store/collect it in the archiving system/software
☰	▾	I collect the title/citation/URL in a document
☰	▾	I bookmark it

\* 10. What type of information sources do you save in your collection? (You may select multiple)

- A copy of the original source material (PDF of an article, saved webpage, etc.)
- A link or Bookmark to the main page of the material
- A citation (Including a bibtex)
- Notes that I have made about my reading of the source material
- Notes about how I found the source, or other meta aspects of the reading
- Other (please specify)

\* 11. After reading or skimming an article, do you take notes? If yes how do you take notes? (You may select multiple)

- I annotate on the digital file
- I print the article and annotate on it
- I take notes on a separate document file
- I don't take notes
- Other (please specify)

\* 12. How do you rank the amount of time you spend on different research tasks? (1- means most time) [You may drag and drop the options]

-   Selecting articles from longer list
-   Storing or collecting articles
-   Cleaning or categorizing collections of articles
-   Reading or skimming articles
-   Taking notes or annotating articles
-   Revising or revisiting the articles from my collection
-   Analyzing or finding patterns in the collected articles
-   Writing the report from my findings



\* 16. To facilitate your research journey, do you use any archiving system/tool? If yes, which of the following apply? (You may select multiple)

- I do not use any archiving system/tool/software
- I save articles on my local file system
- I save articles on a cloud file system
- I use an archiving system/tool/software
- I copy the title/citation/URL and save it in a document file
- I save articles as bookmarks in browser
- I take physical notes on index cards/notebook/sticky notes
- Other (please specify)





\* 22. How much do you agree with the following statement: I generally satisfactorily utilize my collection for writing the final report/paper.

I generally underuse my collection for writing      I optimally use my collection for writing      I am overreliant on my collection for writing

\* 23. How much do you agree with the following statement: I wish I spent more time building my collection.

I should spend a lot less time collecting      I spend the optimal amount of time collecting      I should spend a lot more time collecting

\* 24. How much do you agree with the following statement: I wish I spent more time analyzing my collection.

I should spend a lot less time analyzing      I spend the optimal amount of time analyzing      I should spend a lot more time analyzing

25. Do you have any suggestions for tools, applications or software that could improve the entire process of your research journey?



## Meta-search Information Practices and Support

### Next phase of the study

26. If you are interested to participate in the next phase of this study. Please leave your email, which will be only used to contact you.

End of Survey Study Questions

# Interview Study Script

## Interview Plan & Questions

**Type of Interview:** Semi-structured

**Duration:** 40 minutes

**Medium:** Online

- **UVic Licensed Teams:** To conduct the interview.
- **MS Whiteboard:** For the interactive part of the interview.

### *Table of Contents:*

**Interview Plan & Questions.....1**

A. Consent and Introduction (1 minute):..... 2

B. Main Interview (30-40 minutes):..... 3

    B1. Activity to Capture participants research processes (15 Minutes)..... 3

    B2. Introduction to RaMSeS Model (5 Minutes to explain):..... 4

    B3. Interactive Session Using MS Whiteboard (15 Minutes):..... 5

C. Conclusion and Debrief (1 minute):..... 5

**A. Consent and Introduction (1 minute):**

- Greet the participant.
- Present and explain the consent form, ensuring the participant understands the study's purpose, their role, and their rights (such as confidentiality and the right to withdraw).
- Obtain written consent and Answer any preliminary questions from the participant.

Thank you \_\_\_\_\_ for participating in our study titled "Investigation of Model and Ontology of Literature Research Activities." This study is a part of my thesis. I am Mohammad Shakirul Islam doing my Master's in Computer Science at the University of Victoria, under the supervision of Dr. Miguel Nacenta.

- *If not returned* - I hope you have had a chance to review the consent form I sent over email. It details the purpose of this study, your role, and your rights. Please sign the last page and return it to me before we begin our session.
- ***If returned* - Thank you for signing and returning the consent form.**

For your information, You may withdraw from the study at any time without any consequences or need for an explanation.

Please note that this session will be recorded. If you have no objections, I will start recording from now.

Please let me know if you have any questions; otherwise, we may begin our interview.

## **B. Main Interview (30-40 minutes):**

### **B1. Activity to Capture participants research processes (15 Minutes).**

- **Begin with general questions about the participant's literature processes.**
- **How they teach the literature processes.**

We are researchers interested in understanding how individuals manage their research literature processes and the different phases therein. We also focus on how they compose their RELATED WORKS/LITERATURE REVIEW collections for their projects and research. Our aim is to capture the entire research process to better understand human behavior and discover insights that can help in DESIGNING new software and methodologies or improving existing ones.

The first part of this study aims to capture an overview of the methods you use to teach the process of literature review. You may draw/sketch and describe the process at the same time. I will share a URL for a Microsoft Whiteboard space over chat for you to use. [\[UVic MS whiteboard URL with EDIT ACCESS\]](#)

#### **Prompt:**

You may describe the methods and approaches you use to teach students about literature review processes in academic research. The strategies you do emphasize to your students for effectively searching and identifying relevant literature in their field of study/research project. To capture different aspects I will ask you a few questions.

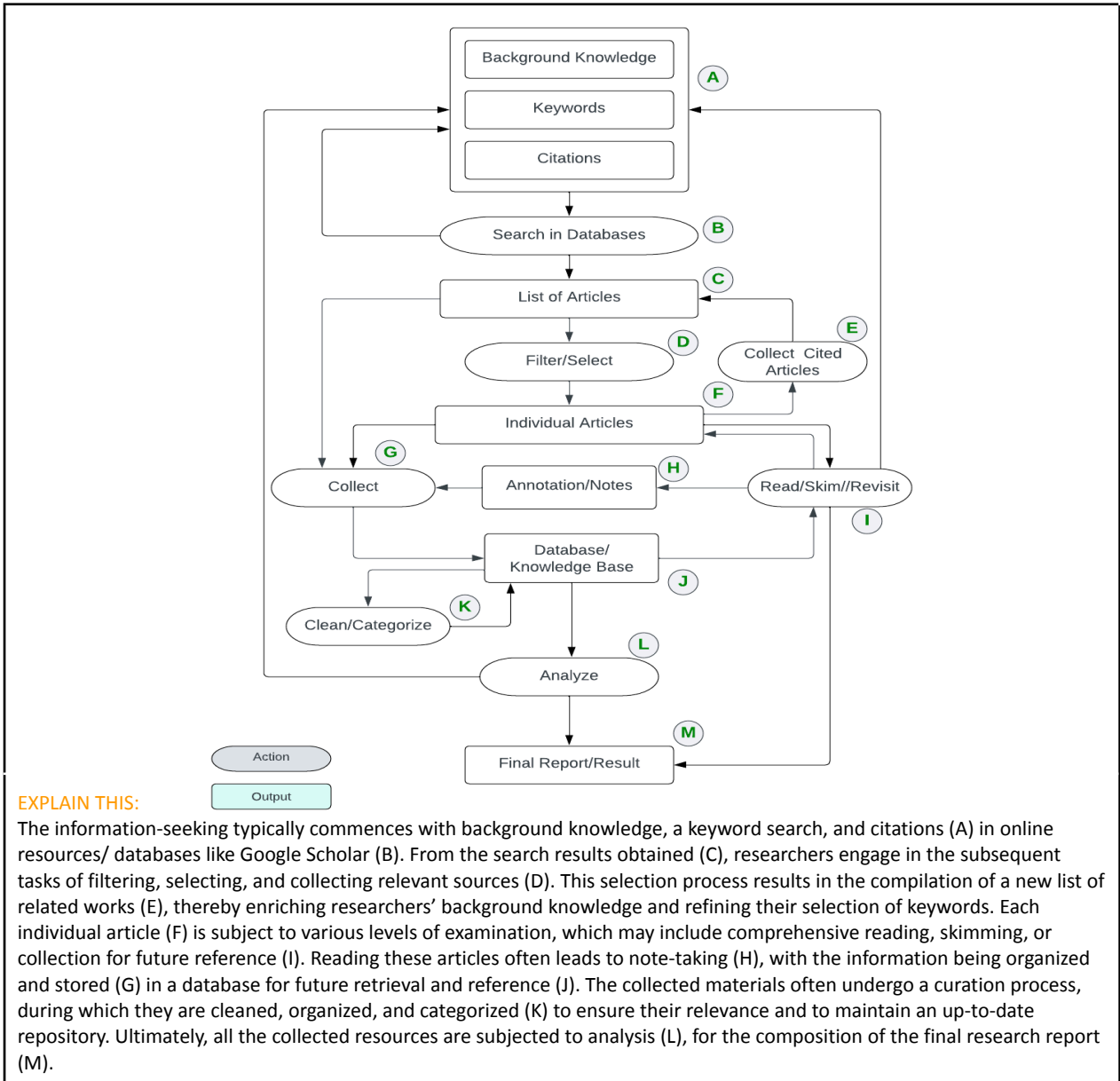
#### **Questions within activity**

1. Can you first explain the different phases in the entire process, from IDEATION to WRITING the RELATED WORK section?
2. Do these phases you mentioned always occur in the same order?
3. Which part of the process do you FOCUS on the most?
4. Which parts do students or researchers handle effectively?
5. Which aspects of the process are MOST challenging?
6. Is the description you've given me related to a specific model or ontology that you've encountered elsewhere?
7. What techniques or tools do you recommend to students for organizing and managing their literature findings?

## B2. Introduction to RaMSeS Model (5 Minutes to explain):

I'm excited to share with you our proposed model, titled 'Researcher's Management of Search and Storage (RaMSeS).' This model outlines the various Actions and Artifacts (or Outputs) involved in the entire research process. We've developed it over several months through numerous iterations leading to this final version.

Let me walk you through it, please feel free to interrupt me if you want to ask any questions.



### **B3. Interactive Session Using MS Whiteboard (15 Minutes):**

- Ask the participant to interact with a visual representation of the RaMSeS model on MS Whiteboard.
- Encourage them to annotate, comment, or suggest modifications to the model.

If you have any questions about understanding this model please ask me [Pause 5 seconds]. I will ask you a few questions related to this model and the process you follow and teach your students.

#### Questions after B2

8. How closely does our model mirror the methods and strategies you've explained earlier?
9. Does our model accurately capture all aspects of your process, or are there elements included that do not reflect your actual processes?
10. How do you think a structured model like RaMSeS could influence or enhance the way you teach research methodology and processes to students?
11. Based on your experience, what improvements or additions could make the RaMSeS model more effective as a teaching tool for the research process?
12. What are your overall thoughts on this model?

### **C. Conclusion and Debrief (1 minute):**

Thank you again for your time and contribution. During our meeting today, we discussed how experts in research literature process different Models of Literature Research Activities for developing their research methodologies. If you want to know a summary of the findings or the study's overall results, it will be released after I finally defended my thesis. It will be available at UVicSpace under the Faculty of Engineering and Department of Computer Science.

Do you have any questions for me before we close this session? If you have further questions later on, please feel free to contact me via email.

I will send you a SurveyMonkey link via email to collect completely optional demographic information. Also, please share your e-transfer email or phone number over email for compensation. ***Have a great day!***

# End of Interview Study Script

# Interview Study Codebook

## Investigation of RaMSeS Model

### Codes\\Codes

Name	Description	Files	References
01.01-05. Participants' Literature Review Process		5	88
01. Literature Phases	Different stages of the literature review process.	5	60
02. Phases Order	The sequence in which the phases occur.	2	3
03. Focus Areas	Parts of the process that receive the most attention	4	5
04. Effective Handling	Aspects of the process that participants manage well.	5	5
05. Challenges	Difficulties encountered during the literature review process.	5	9
01.02. Research Tools and Techniques		5	48
Background Knowledge	How participants build their foundational knowledge.	3	9
Databases	Use and selection of research databases.	4	10
Filtering and Selection	Methods for filtering and selecting relevant literature.	1	2
Keywords	Importance and use of keywords in the research process.	4	8
Research Questions	Formulating and refining research questions.	4	5
Search Techniques	Strategies for searching literature.	5	23
06. Models and Frameworks	Interview Q6: Is the description you've given me related to a specific model or ontology that you've encountered elsewhere?	5	19

Name	Description	Files	References
ACRL Framework for Information Literacy		3	6
Anne Lamott's concept of 'shitty first drafts'		1	1
Big6 Skills of Information Literacy		1	1
Bloom's taxonomy		1	1
Personal Approach from Experience	Individual methodologies based on personal experience.	3	4
The Metamorphosis by Franz Kafka		1	1
07.01. Tools Recommendation in Literature Management	Tools recommended to Students or Literature resources management	5	20
Citation Managers - EndNote, Mendeley, or Zotero		3	5
Non-Users of Citation Managers	Participants who do not use citation management tools.	1	1
Pen-and-Paper	Traditional methods for managing literature.	1	2
Simple Tools - Word or Excel	Recommendations for basic tools like Word or Excel.	2	2
07.02. Teaching		4	24
Citation Management		3	3
Generative AI		1	3
Method Teaching		4	9
Research Management		4	6

Name	Description	Files	References
Sensemaking		3	5
08-09. Evaluation of RaMSeS Model	RQ-1. Does the RaMSeS model comprehensively represent the key activities and sequences undertaken by literature reviewers in academic research?	5	8
Close Representation	How closely the model represents actual processes.	3	12
Comprehensive Model	Feedback on the comprehensiveness of the model	2	5
Thorough Model	Comments on the thoroughness of the model	1	3
10. RaMSeS Enhancing Teaching Research Methodologies	RQ-2. To what extent do experts in research literature processes find the RaMSeS model descriptive and useful for developing their research methodologies?	5	11
Core Emphasis		1	2
Helpful Guide	Utility of the model as a guide.	4	8
Overarching Model	The model's overall scope and utility.	3	5
11. RaMSeS Scope of Improvements	Q-4. What are the primary limitations or missing aspects in the RaMSeS model as identified by experts in research literature processes?	5	26
Gen AI capabilities		1	4
Improvement suggestions		4	18
Layout to a story		1	2
Structured searching		1	3

## End of Interview Study Codebook

# Post-interview Demographics Survey



## Investigation of Model and Ontology of Literature Research Activities

### Demographic Questions

\* 1. What is the highest level of education you have completed or the highest degree you have received?

- Undergraduate student
- Bachelor degree
- Master's student
- Master's level graduate degree
- Doctoral student
- Doctorate level graduate degree
- Post-graduate studies or appointments
- Other (please specify)

\* 2. Which of the following best describes your current occupation?

- Grad Student
- Professor
- Researcher
- Postdoctoral researcher
- Librarian

Other (please specify)

\* 3. What is your domain?

End of Post-interview Demographics Survey