INVESTIGATING THE INFLUENCE OF SUBGOALS ON LEARNING DURING SEARCH

Kelsey Urgo

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Information and Library Science.

Chapel Hill
2023

Approved by:

Jaime Arguello

Rob Capra

Jeff Greene

Gary Marchionini

Heather O'Brien

ABSTRACT

Kelsey Urgo: Investigating the Influence of Subgoals on Learning During Search
(Under the direction of Jaime Arguello)

Search-as-learning research has emphasized the need to better support searchers when learning about complex topics online. Prior work in the learning sciences has shown that effective self-regulated learning (SRL), in which goals are a central function, is critical to improving learning outcomes. This dissertation investigates the influence of subgoals on learning during search. Two conditions were investigated: Subgoals and NoSubgoals. In the Subgoals condition, a tool called the Subgoal Manager was used to help searchers to develop specific subgoals associated with an overall learning-oriented search task. The influence of subgoals is explored along four dimensions: (1) learning outcomes; (2) searcher perceptions; (3) search behaviors; and (4) SRL processes. Learning outcomes were measured with two assessments, an established multiple-choice conceptual knowledge test and an open-ended summary of learning. Learning assessments were administered immediately after search and one week after search to capture learning retention. A qualitative analysis was conducted to identify the percentage of true statements on open-ended learning assessments. A think-aloud protocol was used to capture SRL processes. A second qualitative analysis was conducted to categorize SRL processes from think-aloud comments and behaviors during the search session. Findings from the dissertation suggest that subgoals improved learning during search. Additionally, it seems that subgoals helped participants to better retain what was learned one week later. Findings also suggest that SRL processes of participants in the Subgoals condition were more frequent and more diverse. SRL processes that were explicitly supported by the Subgoal Manager seemed to be more frequent in the Subgoals condition as well as SRL processes that were not explicitly supported.

To my husband, John Hutchens, our baby, Dashiell Urgo-Hutchens, and my parents, Cynthia Crowe-Urgo and Michael Urgo. I love you all so much.

ACKNOWLEDGEMENTS

I would like to thank my friend, partner, and husband, John. Thank you for explaining complex mathematical ideas, cooking meals, taking walks, and telling absurd jokes. It is because of your humor, brilliance, support, and love that I made it through to the finish line. I would like to thank my baby, Dashiell, for bringing tranquility, delight, curiosity, and fun. You are extraordinary. I would like to thank my parents, Cynthia and Michael, for always being available to talk and listen and for supporting me in all my wild endeavors. I would like to thank my brother, Creighton Urgo, for his motivation and insight. I would like to thank my aunts, Kathleen Crowe and Kristine Zavoli, and uncle, Walter Zavoli, for their encouragement and unrelenting belief in my abilities. I would like to thank Emmy and Ruth for giving me balance and perspective.

I am grateful for my profoundly supportive and encouraging committee members: Jaime Arguello, Rob Capra, Jeff Greene, Gary Marchionini, and Heather O'Brien. Throughout the PhD process, they each provided perceptive questions and important feedback that strengthened my dissertation work.

I would like to thank my advisor, Jaime Arguello. Thank you for your availability throughout the years-long PhD process, participating in weekly (or more) meetings and answering countless emails. Thank you for taking the time to think through difficult research problems with me, edit pages (upon pages) of my writing, and for helping me to process rejections and to understand that they are a useful part of the academic process. Thank you for listening to and respecting my ideas. You have made me a better writer and a considered, confident researcher. I am so grateful for your guidance.

I would like to thank Rob Capra. Thank you for providing a collaborative and positive research environment. Thank you for all the lunches, research discussions, and thoughtful, motivating feedback you have provided throughout the years. Thank you for valuing and considering my ideas in an authentic way. I am grateful to you for making a productive and supportive space for students to learn and grow.

I would like to thank Gary Marchionini. Gary was the first person I met at SILS and he is one of the main reasons why I applied for, and accepted the offer to embark upon, the PhD program. Thank you for your enthusiasm and big-picture thinking that have helped me to achieve the career of which I once only dreamed.

I would like to thank Jeff Greene. I am thankful to Jeff for introducing me to the world of self-regulated learning. Thank you for taking the time to teach an independent study course with me. Thank you for helping me to bridge my work into another discipline in a clear and focused way.

I would like to thank Heather O'Brien. Thank you for your unending kindness and enthusiasm for my work. I am so grateful to you for believing in me, taking time to meet, and for providing feedback that strengthened my work and helped me to see new directions in my own future research.

I would like to thank Yue Ray Wang. I was fortunate enough to have Ray for two courses at SILS, Information Retrieval and Data Mining. Thank you for your optimism and willingness to collaborate and share ideas.

I would like to thank my friends. I am thankful for Felicia Simpson for sending supportive messages, having lunches, and for understanding what it takes to be an academic and a mother. I am grateful to Terrell Russell and Stephanie Suber for their ongoing support and confidence in my abilities. I am also thankful for my education and the careers that prepared me for, and led me to, the PhD program at SILS. I would like to thank the Renaissance Computing Institute. I am grateful for the incredible, intelligent people of RENCI and, in particular, the support of Stan Ahalt, Jay Aikat, and Ray Idaszak.

I am grateful for the collaboration, levity, and kindness of my fellow (and former) SILS and Interactive Information Systems Lab students: Sandeep Avula, Bogeum Choi, Anita Crescenzi, Yuan Li, Austin Ward, and Yinglong Zhang. Thank you for the lunches, coffees, Zoom meetings, and impromptu chats that provided me with advice, encouragement, and motivation to complete a challenging and enlightening journey.

Thank you to those who were pivotal in carrying out my dissertation study. I would like to thank Antoinette Odendaal and Abraham Tucker for their support, time, and contribution of their expert knowledge of biology. I would like to thank the organizations that provided funding to support my

dissertation work [1]. I would also like to thank the participants in my dissertation study. Thank you to SILS staff members Lara Bailey, Kaitlyn Marlowe, Susan Sylvester, and Michelle Taylor who helped me successfully navigate both the execution of my dissertation study and the PhD program at SILS. I am so grateful to you all.

[1] My dissertation study was directly supported by the Edward Holley Research Fund and Carnegie Grant from the School of Information and Library Science in the University of North Carolina at Chapel Hill. The research described in this dissertation was also partially supported by NSF grant IIS-2106334.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

Search has become a part of our daily lives. Whether we aim to find an address to a doctor's office or understand the forces behind Bernoulli's principle, we turn to a search engine for help. In response to these needs, search engines require a user to translate an information need into a single query, then scan through a list of retrieved links. This search paradigm has proven successful for simple lookup tasks. However, such a model leads to struggle and frustration when searchers are faced with complex learning-oriented search tasks. Cognitive processes that require relating concepts, evaluating the effectiveness of procedures, and constructing new knowledge are arduous and at times impossible with typical search systems. In light of current systems' dysfunction, we *need* systems to assist us in complex learning processes. Rieh et al. [4] summarized findings from Eickhoff [5] and Bailey [6] that found 20% of search tasks are categorized as complex and multi-step (features indicative of learning-oriented search tasks). Multiple seminars have specifically addressed the need to better support learning during the search process [7, 8, 9, 10]. The community involved in these seminars coined the term search-as-learning to describe this work.

Early work in information retrieval (IR) focused on the need for precision-driven search systems in service of factoid or lookup tasks. More recently, the IR community has recognized the need to support complex learning tasks. In response, IR researchers established the search-as-learning community. Search-as-learning was established primarily through The Second Strategic Workshop on Information Retrieval in Lorne (SWIRL) [8] and Dagstuhl Seminar 17092 [7]. Several core goals of search-as-learning were identified in these meetings—(1) understanding search as a learning process; (2) measuring learning outcomes during search; (3) understanding the contexts in which people search to learn; and (4) developing tools to support learning during search. Prior work in search-as-learning has investigated the role of the individual [11, 12, 13, 14], the task [15, 16, 17], and the system [18, 19, 20, 21, 22, 23] on learning during search. Additionally, prior work has identified search behaviors that may be predictive of learning during search [24, 25, 26, 27, 28, 29, 30, 31]. My dissertation bridges three areas of prior work: (1) search-as-learning; (2) self-regulated learning; and (3) goal-setting.

Researchers in the learning sciences have studied the critical role of effective self-regulated learning (SRL) in improving learning outcomes [32, 33, 2, 34, 35]. SRL is demonstrated by an active, reflective learner that effectively monitors and controls their own learning [36]. SRL has been categorized into several macro-SRL processes that contain specific micro-SRL processes [37]. Macro-SRL processes include *Planning*, conceiving an approach to achieving an overall learning task (e.g., goal-setting), *Strategy Use*, selecting strategies for achieving an overall learning task (e.g., summarization), and *Monitoring*, tracking progress by comparing what has been learned with standards and qualities of the ideal target (e.g., monitor progress toward goals) [38]. Goal-setting is particularly important to SRL. First, goals prompt learners to reflect on their understanding of the task. Second, goals focus attention on planning and influence strategy choice for goal achievement. Third, goals provide standards for monitoring progress [39]. Goals are at the core of effective SRL, acting as a source of external feedback enabling learners to monitor their progress so that they may respond in a way that positively impacts goal attainment (e.g., working harder, changing strategies). SRL *should* be central to search-as-learning work as effective SRL is crucial to improving learning outcomes. Within SRL, goal-setting is of critical interest to search-as-learning as it serves a central function in effective SRL and is one of the *most* important factors in predicting learning outcomes [33]. To investigate the role of goal-setting in learning during search, I have developed a tool called the Subgoal Manager. The Subgoal Manager is designed to *encourage* and *facilitate* goal-setting and prompt *monitoring* of goal progress.

1.1 The Subgoal Manager

The Subgoal Manager is a tool that allows searchers to externalize and monitor their progress of learning-oriented subgoals. The Subogal Manager is shown in Figure 1.1. The Subgoal Manager allows searchers to develop, write, and modify subgoals, monitor progress toward subgoals, and add information associated with subgoals while searching. Additionally, the Subgoal Manager provides tool tips that specify four qualities of good subgoals (i.e., specific action, information, criteria, and time [details are discussed in Chapter 2]) that encourage and support searchers in developing more achievable subgoals. The Subgoal Manager allows searchers to add or delete subgoals throughout the search process. Subgoals can be deleted with a grey x. As an assurance check, a pop-up asks searchers if they are sure they want to delete the subgoal before the subgoal is actually deleted. Subgoals can be added by clicking the "Add Subgoal" button at the bottom of the Subgoal Manager. Each subgoal is collapsed by default to minimize clutter and can be expanded by clicking the blue caret on the right. Clicking this caret reveals text editors associated with each subgoal. Each text editor has functionality one might expect from such an editor, bolding, highlighting, underlining, and italicizing text, along with other formatting options such as bulleted lists and headings within the text. The editors allow searchers to take notes and monitor their progress toward achieving each subgoal. Finally, the Subgoal Manager provides a "subgoal complete" checkbox in each subgoal. When the subgoal complete checkbox is checked, the subgoal collapses hiding the text editor and the subgoal darkens in color to a deeper shade of grey. This functionality further supports subgoal monitoring. The presence of the subgoal complete checkbox prompts searchers to evaluate if they have achieved the subgoal and decide if the subgoal has been completed.

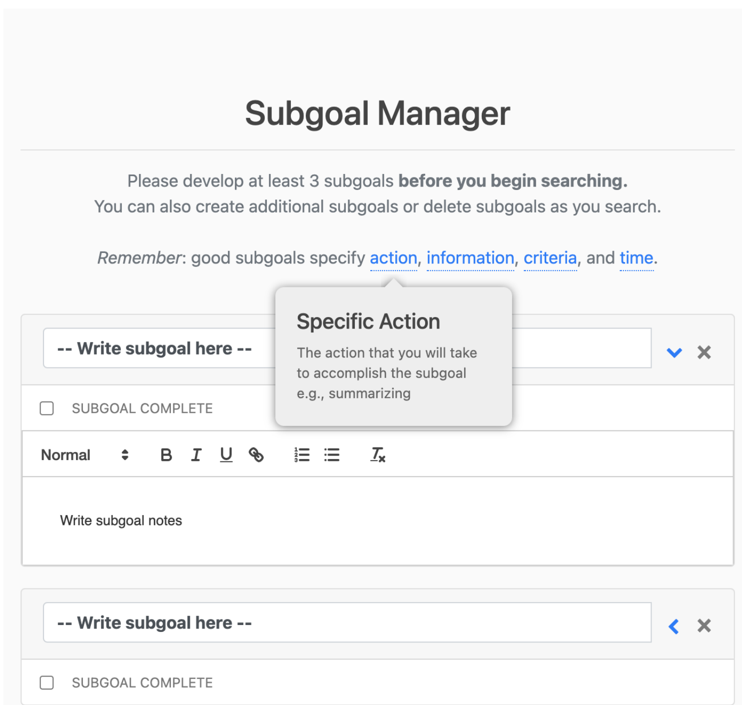

Figure 1.1 The Subgoal Manager

The Subgoal Manager is designed to support goal-setting and SRL processes. Specifically, the Subgoal Manager allows the searcher to monitor their goal progress and adapt their strategies based on the evaluation of progress. For example, a searcher may have set the following subgoal: *In 10 minutes, explain the relationship between diffusion and osmosis, include 3 similarities and 3 differences.* After reading this subgoal, the learner may be prompted to monitor their progress toward the subgoal. Perhaps they notice they have written 3 similarities in the associated text editor but not identified any differences between the concepts. Realizing this discrepancy may encourage the learner to adapt their approach and select the strategy of summarization. In reaction, the learner decides to summarize everything they read on the current web page relating to each concept in the subgoal text editor. Through summarizing, the learner is able to add two differences between the concepts. The Subgoal Manager was used in both a preliminary study (discussed next) and my dissertation study.

1.2 Preliminary Study

Broadly speaking, my dissertation study will investigate the role of subgoals during learning-oriented search. My dissertation research builds upon a preliminary study that also investigated the role of goal-setting on learning during search. In the preliminary study, we investigated the influences of a tool that encouraged and supported the deconstruction of learning-oriented search tasks into specific subgoals (assigned or self-set). Participants were assigned to one of three conditions in which participants either—(1) developed their own subgoals; (2) followed pre-established subgoals; or (3) were not instructed to create subgoals nor were assigned subgoals. Participants who set their own subgoals perceived higher level of support for SRL processes of *Planning*, *Evaluating Progress*, and *Adapting* (i.e., changing strategy or approach) during the search task (Appendix A provides details of this preliminary study). Additionally, participants who set their own goals had the highest learning outcomes. Because participants who set their own subgoals had the highest learning outcomes and perceived greater SRL support, my dissertation study focuses only on differences between the self-set subgoals and no subgoals conditions.

1.3 Experimental Conditions

My dissertation study investigated two experimental conditions: Subgoals and NoSubgoals.

Subgoals: In this condition, participants were provided with the Subgoal Manager (Figure 1.1). In a video, participants were provided with an overview of the functionality of the Subgoal Manager (e.g., expanding text editor underneath subgoals to take notes, marking subgoals complete). Participants were informed that subgoals can be modified, deleted, and added throughout the search process. Additionally, in the video, participants were introduced to ideal subgoal characteristics (i.e., specific time, action, standard, and content) and were advised to develop their own subgoals with these ideal characteristics. Before beginning their search, participants were asked to develop at least three subgoals. In short, this condition introduced participants to ideal subgoal qualities and asked participants to develop at least three subgoals using the Subgoal Manager before beginning the search task.

NoSubgoals: In this condition, participants were provided with the Text Editor similar to a Word doc. In a video, participants were provided with an overview of the functionality of the Text Editor (e.g., italicizing words, creating headings). Participants were not prompted to write subgoals.

1.4 Research Questions

My dissertation research builds on my preliminary work and explores the role of learning-oriented subgoals to improve learning during search. My interest in the role of goal-setting during search is motivated by three trends observed in prior work. First, an extensive body of work has shown that effective goal-setting serves a critical role in SRL and improving learning outcomes. Second, in my preliminary study (Appendix A), participants achieved better learning outcomes when given access to the Subgoal Manager *and* were able to develop their own subgoals (i.e., self-set vs. assigned). Third, in my preliminary study (Appendix A), participants also reported that the Subgoal Manager provided greater support for SRL. My dissertation research will investigate the four following research questions:

- **RQ1**: How does the subgoal condition influence learning during search?
    - *RQ1-A*: On post-task learning scores?
    - *RQ1-B*: On delayed learning scores (i.e., retention)?
- **RQ2**: How does the subgoal condition influence searcher perceptions? I investigate perceptions of interest, knowledge increase, satisfaction, difficulty, SRL support, and engagement.
- **RQ3**: How does the subgoal condition influence search behavior (e.g., queries, clicks, etc.) during search?
- **RQ4**: How does the subgoal condition influence SRL processes? I investigate *Planning* (e.g.,

subgoals, prior knowledge activation, and recycle goal in working memory), *Strategy Use* (e.g., taking notes, reading notes, and comparing & contrasting), *Monitoring* (e.g., monitoring progress toward subgoals, content evaluation, and judgment of learning), and *Interest*.

In Chapter 2, I provide an overview of the three major fields my dissertation builds upon—(1) search-as-learning; (2) self-regulated learning; and (3) goal-setting. Additionally, I review relevant computer-based learning environments. Finally, because the measurement of learning plays an integral role in my dissertation, I discuss learning assessments used both within and outside of search-as-learning. In Chapter 3, I discuss my four research questions. Some research questions are accompanied with hypotheses about the expected effects of the subgoal condition. Other research questions are investigated in an exploratory manner. For each hypothesis, I describe prior work that informs the expected effects of the subgoal condition. In Chapter 4, I review the methodology for my dissertation study. This includes details of the study protocol, participant recruitment, instructional videos, learning-oriented search task, search system, questionnaires, learning assessments, think-aloud and search behavior data, and data analysis for each research question. In Chapter 5, I present results from all four **RQ**s. In Chapter 6, I discuss the findings and implications from this dissertation. In this chapter, I also discuss directions for future work that stem from results. Finally, I conclude in Chapter 7.

CHAPTER 2

Background

My dissertation builds on three major fields of study—(1) search-as-learning; (2) self-regulated learning; and (3) goal-setting. In this chapter, I provide a brief overview of each and highlight studies that directly motivate my dissertation study. Additionally, because my dissertation study explores support for learning during search using a novel auxiliary search tool, I review prior studies involving similar computer-based learning environments. Finally, because measurements of learning play a critical role in search-as-learning and my dissertation study, I provide an overview of learning assessments used in prior work. In this respect, I review learning assessments used in previous studies in search-as-learning and educational research.

2.1 Search-as-Learning

Learning has been a subject of research in IR for many years and there are two meetings central to the establishment of the search-as-learning community: The Second Strategic Workshop on Information Retrieval in Lorne (SWIRL) [8] and Dagstuhl Seminar 17092 [7]. In 2012, the three-day SWIRL workshop emphasized the importance of supporting learning during search as one of several emerging topics. In 2017, Dagstuhl was entirely dedicated to the topic of search-as-learning. Discussions from the seminar established four main research issues in search-as-learning—(1) examining search as a learning process; (2) measuring learning performance and outcomes during search; (3) investigating the contexts in which people search to learn; and (4) developing tools and interventions to promote learning during search.

In order to design search systems that better support learning, we must first understand the factors or characteristics that affect learning during search. The majority of search-as-learning work has focused on characteristics that impact learning during search from three perspectives—(1) characteristics of the individual [12, 11, 13]; (2) characteristics of the task [40, 41, 42, 43]; and (3) characteristics of the system [44, 19, 18, 20]. Prior work in search-as-learning has also aimed at understanding how specific search behaviors can predict learning outcomes [24, 25, 26, 27, 28, 29, 30, 31].

2.1.1 Influence of Individual Characteristics on Learning During Search

A set of prior work has focused on the the effect of individual characteristics on learning during search [11, 12, 13, 14]. Willoughby et al. [11], O'Brien [12], and Roy et al. [13] explored the role of domain knowledge on learning during search. Willoughby et al. [11] asked participants to write summaries of what they knew in a particular domain. Participants were grouped into different conditions, with some participants searching for 30 minutes before completing summaries while others produced summaries without searching. Those participants who searched wrote summaries with more accurate facts. The researchers also found this effect to be stronger for participants with greater prior knowledge, hypothesizing that these participants were able to search more effectively. O'Brien et al. [12] asked participants to complete knowledge summaries before and after learning-oriented search sessions. When compared with domain experts, novices had larger improvements in their summary scores before and after search. This may be explained by novices uncovering a greater amount of *new* information during search. Roy et al. [13] asked participants to complete knowledge assessments intermittently throughout the search session. Domain knowledge affected *when* participants had greater knowledge gains. Novices learned more at the beginning of the search session while experts learned more toward the end.

Distinct from domain knowledge, Pardi et al. [14] explored the role of both working memory capacity and reading comprehension on learning during search. Participants were asked to write knowledge summaries before and after searching. Both working memory and reading comprehension had positive effects on learning, reflected in the number of relevant concepts included in participant summaries.

2.1.2 Influence of Task Characteristics on Learning During Search

Prior work has explored the influence of task complexity on learning during search [15, 16, 17]. In these studies, Ghosh et al. [15], Kalyani and Gadiraju [16], and Liu et al. [17] all leveraged the cognitive process dimension from the A&K taxonomy [45] to vary search tasks by complexity. Ghosh et al. [15] found that participants had significant knowledge gains across *all* tasks. However, participants had *smaller* knowledge gains for tasks that were more complex. Kalyani and Gadiraju [16] asked participants to complete tasks associated with all six cognitive processes from the A&K taxonomy. Similar to Ghosh et al., knowledge gains were lower in more complex tasks (i.e., apply $<$ analyze).

Liu et al. [17] asked participants to complete both a receptive task and a critical task. Receptive tasks were simpler (i.e., remember or understand) while critical tasks were more complex (i.e., evaluate). Participants were asked to develop mind maps before and during the search session. These were used to analyze their evolving knowledge. During receptive tasks, participants made edits to their mind maps throughout the *entire* search session. Conversely, during critical tasks, participants made more edits to their mind maps toward the *end* of the search session.

Apart from task complexity, prior work has examined the influence of task session on learning during search. Liu et al. [46] asked participants to complete three subtasks relating to one overall topic. Liu et al. developed two distinct sets of subtasks, those that built upon one another (dependent subtasks) and those that were independent of one another (parallel subtasks). Participants reported higher levels of topic familiarity after completing each subsequent subtask. However, participant topic familiarity leveled off more quickly in the parallel subtask condition (versus dependent subtask). While this study did not investigate the role of goal-setting during search, this result suggests that participants achieved better learning outcomes when they pursued subgoals that built on one another.

2.1.3 Influence of System Characteristics on Learning During Search

Prior work has investigated how certain tools and interface features can affect learning during search [18, 19, 20, 21]. Freund et al. [19] explored the effect of two factors: (1) plain text layout versus HTML layout with distracting elements (e.g., ads); and (2) with versus without a sticky notes tool that allowed participant to annotate articles. Participants had better reading comprehension outcomes with the plain text vs. HTML condition. However, when provided with the sticky notes tool, learning outcomes were similar in both plain text and HTML conditions. Kammerer et al. [18] investigated the impact of an exploratory search system that also afforded the ability to filter search results with social tags. Participants had higher learning outcomes in the experimental system with social tag filtering. Roy et al. [20] examined the effect of a search system that allowed participants to highlight text and make notes. Participants who used the highlighting tool had summaries with more subtopics, while those that used the note-taking tool had summaries with a greater number of facts. When provided *both* tools, participants did not have increased learning outcomes. The authors conjectured that access to both tools led to cognitive overload. Câmara et al. [21] investigated the influence of interface features that displayed domain subtopics and also displayed the participant's coverage of subtopics across a search session. These features did not increase learning outcomes.

Participants viewed more search results more quickly, exploring more subtopics superficially. This may indicate that some search features can unintentionally cause searchers to follow strategies that reduce *depth* of learning.

Studies have also explored the influence of the retrieval algorithm on learning during search. Syed and Collins-Thompson [22] developed a retrieval algorithm that favored documents with a greater density of vocabulary words that participants were asked to learn. Participants had higher learning outcomes using a search system with this experimental algorithm when compared to a baseline search system. Weingart and Eickhoff [23] investigated the influence of a variety of established retrieval techniques on learning during search. When compared with document retrieval, *passage* retrieval had a positive effect on learning. Query expansion, however, had a negative effect on learning.

2.1.4 Behaviors Predictive of Learning During Search

Much prior work has also explored search behaviors that may be predictive of learning during search [24, 25, 26, 27, 28, 29, 30, 31]. Studies have found that searchers with better learning outcomes have a tendency to: (1) spend more time reading documents [25, 24, 26, 28, 22]; (2) issue queries with more advanced or uncommon vocabulary [25, 24, 30]; (3) issue more diverse queries within the session [31]; (4) review more search results that are relevant and novel [25, 29]; and (5) visit sources that are more suitable to the task, such as encyclopedic sources during *receptive* tasks and Q&A sources during *critical* tasks [27].

It is important to understand the relation between specific search behaviors and learning. That being said, search behaviors can be ambiguous with respect to learning. For example, studies have found that searchers achieve better learning outcomes when they spend more time on landing pages [25, 24, 26, 28, 22]. However, "time on landing pages" is a coarse measure. For example, Bhattacharya et al. [30] conducted an eye-tracking study to investigate the relation between reading behaviors and learning. Participants with more eye-regressions (i.e., evidence of *re-reading*) had *lower* learning outcomes. This result suggests that a measure such as "time on landing pages" can be ambiguous. It may predict positive knowledge gains if a searcher is successfully extracting new knowledge while reading or negative knowledge gains if a searcher is struggling to understand information.

2.2 Self-Regulated Learning

For decades, researchers in the learning sciences have studied the components and processes of self-regulated learning (SRL) and the critical role of effective SRL in learning achievement [32, 33, 2, 34, 35]. Prior work had found that SRL processes play a crucial role in learning [47, 48, 49, 50]. SRL is historically rooted in related but distinct areas of psychological and educational research (i.e., cognition/metacognition, motivation, behavioral control, and developmental processes) [35]. Two symposia [51, 52] established a unified definition of SRL that included metacognition, motivation, and behavior as key components in the learning process. This unified definition situated students as proactive and strategic learners in contrast with earlier work that portrayed them as passive learners in their environment [53]. SRL is a reflective and active process in which a learner monitors and controls their own learning in order to achieve desired learning goals [54, 35, 55]. Through goal-setting, self-regulated learners generate feedback loops that allow for monitoring of progress and enacting of strategies when tactics are not producing the desired learning outcomes.

2.2.1 Winne & Hadwin Model of Self-Regulated Learning

Several models of SRL emerged from a set of prior work that investigated how learners engage in complex tasks [38, 56, 57, 58, 32]. Each of these models utilize a slightly different lens [59, 33]. Zimmerman's model of SRL [57] is rooted in socio-cognitive theory based on the work of Bandura [60], highlighting the social foundations of cognition and behavior. Pintrich's model of SRL [56] also comes from a social cognitive approach with research mainly focused on motivation [59]. Boekaerts's model of SRL [32] was influenced by Action Control Theory [61] with research also focused on student motivation. The Winne and Hadwin model of SRL [2] builds on the work Bandura and Zimmerman, highlighting metacognitive knowledge and skills. Despite these differences, these SRL models share a common theme. Self-regulated learning is generally defined by an active, reflective learner that is effective at monitoring and controlling their own learning [36]. Metacognitive knowledge and skills are critical parts of SRL. Metacognitive knowledge includes a learner's awareness of their own learning as well as knowledge of learning strategies in general. Metacognitive skills include monitoring (identifying matches and mismatches between standards and products of a task) and control (changing and enacting different strategies based on feedback). Though many models of SRL exist, I have selected the Winne and Hadwin model [38] because it is supported by evidence from many empirical studies [36, 62, 63, 64, 65, 66, 67]. I am also using the Winne and Hadwin model because it underscores the importance of goal-setting as a distinct and important phase of effective learning.

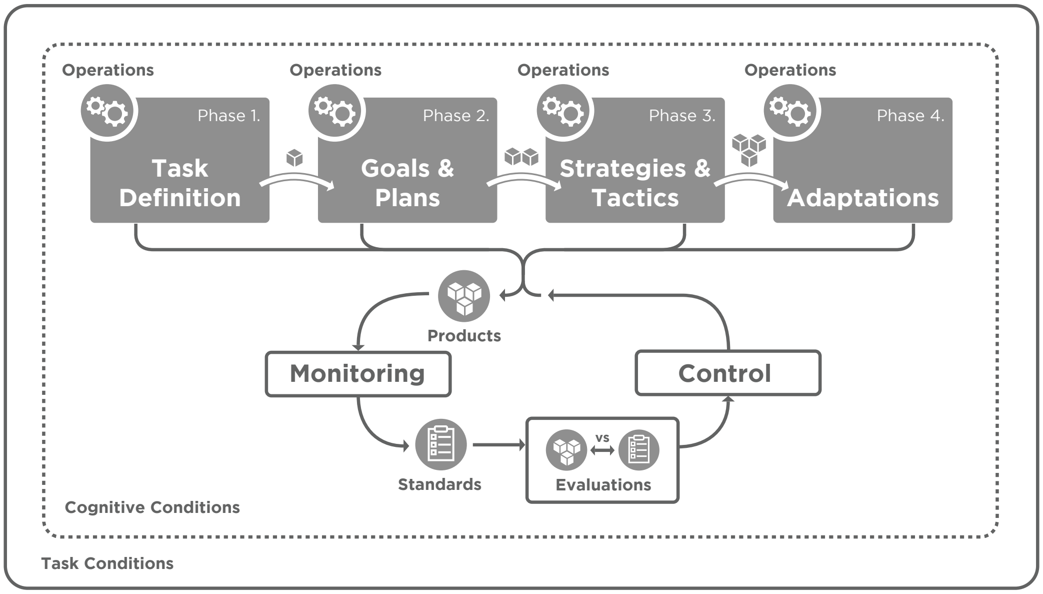

Figure 2.1 Conceptualization of Winne & Hadwin Model

Winne et al. present an iterative model of SRL composed of four weakly sequential phases. Figure 2.1 is my streamlined interpretation of the Winne & Hadwin model. In the first phase, the learner develops a working definition of the learning task. This definition is developed using external and internal resources. External resources may include a task description or a teacher's instruction. Internal resources may include prior knowledge and perceived abilities. The task definition influences all subsequent SRL phases.

In the second phase, the learner sets goals, making a plan to complete the task defined in Phase 1. Learners also choose cognitive tactics and strategies deemed necessary to achieve the goals. These goals are modeled as Standards in Figure 2.1. Goals or standards are dynamic. They may be updated in subsequent cycles as deeper understanding is acquired. For example, a learner might be given an assignment to characterize qualities they believe are shared by a list of famous paintings. Initially, a learner might set a goal of 3 qualities based on the assignment description and typical expectation of the teacher who gave the assignment. However, after some initial research, the learner finds that her resources online about the famous paintings are abundant and updates her standards. She now believes 5 shared qualities would be a more fitting standard given the available resources.

In the third phase, the learner uses the tactics and strategies identified in Phase 2. During this phase, tactics construct information in the learner's working memory. The learner implements activities or studying skills in order to learn the material.

The fourth phase involves learners making adaptations. This is a time of self-reflection when a learner looks back and evaluates the success and/or failure across the previous phases. Based on the feedback from this phase, a learner may identify adaptive or maladaptive behaviors that helped or hindered the learning process. Adaptive behaviors are actions that include adjusting plans, strategies, or task perceptions. Maladaptive behaviors include failing to interact productively, using static strategies when negative judgments of learning have occurred, and ignoring experienced challenges [68]. A learner may also identify strategies that would improve success in the future given the challenges found during this reflection period.

Throughout each phase of the model, a learner uses conditions to inform operations (cognitive processes) and standards (criteria for success). Conditions are the available materials and limitations afforded to a learner for a given task or environment. These conditions are categorized as either cognitive conditions or task conditions. Cognitive conditions include prior knowledge, motivation, and understanding of the task. Task conditions include the physical space, time, and resources. Conditions inform operations and standards throughout each phase of the model. For example, if a learner has little prior knowledge about the task, is in a loud, distracting environment, and pressed for time, their standards might be different than a task with which they have lots of prior knowledge, a quiet workspace, and ample time.

Standards are the criteria a learner has deemed important for optimal completion of a task. The operations are the cognitive processes noted by Winne with the acronym SMART-searching, monitoring, assembling, rehearsing, and translating. There are also higher-order strategies that fall under operations such as note-taking, summarizing, making inferences, or hypothesizing. As these operations create products (or updates to prior phases) learners may enact control if progress deviates from goals. One of these critical operations, metacognitive monitoring is crucial to this process. This metacognitive monitoring of learning coupled with metacognitive control are the "pivots upon which each of the four phases turn." [69, p. 469]

Metacognitive monitoring is the process of comparing what has been learned with standards and qualities of the ideal target. Metacognitive monitoring includes the use of feeling of knowing, judgment of learning, and monitoring of goals [70]. Feeling of knowing is defined by a learner having some understanding but having difficulty with recalling the information. For example, after reading a passage, a learner might think "Oh, I remember this, I think it was called hypertropical...or something like that. I can't quite remember the word." Judgment of learning is defined by a learner becoming aware of what they don't know or understand. For example, after reading a passage, a learner might think "I really don't know what this concept is, this is hard." Monitoring of goals is the assessment of whether a previously set goal has been achieved. Metacognitive monitoring is a special case of monitoring as it is "applied to topics that *are* the *subject matter* the learner is studying." [69, p. 471] This monitoring provides the learner with assessment information to facilitate decisions surrounding metacognitive control.

Metacognitive control is the implementing of tactics and strategies based on the information gained from metacognitive monitoring. Metacognitive control can take the form of allocating more time or attention to more difficult components of a task. The following is an example of metacognitive monitoring and control functioning together: a student realizes that they do not know how to solve a math problem and decides to skip the problem. In this sequence, the student uses metacognitive monitoring, specifically judgment of learning, to identify that a problem is beyond their understanding. In response to this monitoring assessment, the student enacts metacognitive control by implementing a strategy, skipping the problem and moving on to other problems they might be able to answer. As another example: a student reviews their set goal *define Bernoulli's principle in my own words* and, realizing they need more information on Bernoulli's principle, they decide to search online for *Bernoulli's principle definition*. In this example the student engaged in metacognitive monitoring of goal progress and realized the goal had not yet been met. In response, the student enacts metacognitive control by implementing a strategy, issuing a goal-directed search.

Models of SRL underscore the importance of goal-setting in effective learners. Schunk and Greene clearly state the role of goals in SRL, "Self-regulated learners set goals and metacognitively monitor their progress toward them. They respond to their monitoring, as well as to external feedback, in ways they believe will help them attain their goals, such as by working harder or changing their strategies. Goal attainment leads to setting new goals." [71, p. 1] In the next section I explore methods for observing, collecting, and measuring self-regulated learning.

2.2.2 Measuring Self-Regulated Learning

Collecting SRL data can be quite challenging for two main reasons. First, although researchers have developed various self-report inventories to capture SRL processes (e.g., MAI [72], MSLQ [73], and LASSI [74]), prior work has shown that learners are not accurate reporters of their own SRL processing [75]. Second, "Effective SRL processing is dynamic and adaptive, occurring as a series of events over the entirety of a learning task." [76, p. 118] This dynamic nature of SRL means that a dynamic method of data collection is typically necessary to accurately capture SRL processes as they unfold. Greene et al. [77] have proposed a "right tool for the job" approach to SRL when collecting SRL data. For example, while "motivational and dispositional aspects of SRL may be best captured by self-report data [...] more transient, dynamic task-specific aspects may be best captured by TAPs [think-aloud protocols]." [77, p. 323] My dissertation study aims to capture the frequency and types of SRL processes that occur while learning during search. Coarse-grained self-report data of SRL processes, that has been shown to be unreliable, is not sufficient to understand how SRL changes across goal-setting conditions. For this reason, in my dissertation study I investigated think-aloud data and recorded screen activities as a means to uncover the frequencies and diversity of SRL processes that occurred.

Prior work in SRL has coded think-aloud and interaction data to better understand the frequency and types of SRL processes that occurred. Particularly important to my dissertation is the work of Greene et al. that have coded think-aloud data into macro-SRL and micro-SRL processes [78, 37, 79]. Macro-SRL processes include those related to the major components of Winne & Hadwin's model of SRL (e.g., planning, monitoring, strategy use), while micro-SRL processes are categorized within macro-SRL processes (e.g., developing subgoals is a micro-SRL process of the macro-process planning). Researchers have successfully implemented think-aloud protocols to better understand the types of SRL processes that support learning in computer-based learning environments. For example, Azevedo [80] has led a body of work that has investigated SRL processes during science learning in a computer-based learning environment. "Findings indicated that conceptual understanding in science was more likely when participants used high-level learning strategies (e.g., knowledge elaboration) as opposed to low-level ones (e.g., rereading [...]), activated prior knowledge and established subgoals relevant to their learning, planned their time and effort carefully, and frequently monitored their growing understanding." [77, p. 331]

2.3 Goal-Setting

Goal-setting is identified across SRL models as an important part of the self-regulation process. Goals are particularly central to the Winne and Hadwin model of SRL, having three main functions. First, goals prompt learners to consider their task understanding. Second, goals direct attention toward planning and affect strategy choice for achievement. Third, goals provide standards for monitoring and evaluating progress [39]. Goals are the aim or purpose of a learner and can be long or short term. Prior work has shown that goal-setting leads to increased learning outcomes. Goal-setting is integral to effective SRL and there is substantial empirical research that supports the impact of goal-setting on increased learning outcomes. This strong evidence suggests goal-setting should be integrated into search system design in order to better support learning during search.

Goal-setting theory was developed in psychology in hopes to better understand the notion of motivation. Locke and Latham were central to this effort, contributing decades of work on the subject. "A goal is the object of aim of an action, for example, to attain a specific standard of proficiency, usually within a specified time limit." [81, p. 705] In the SRL process, these goals act as long and short term sources of feedback in metacognitive monitoring and control.

Goals are defined by a learner's purpose and are characterized by quantity, quality, and performance (attainment) [82]. Goal-setting is the creation of an objective that is the aim of the learner's actions [55]. Locke and Latham have invested over half a century investigating goal-setting [83]. Their analysis of hundreds of studies point to several important goal characteristics that affect achievement: (1) difficulty of the goal; (2) specificity of the goal; (3) proximal versus distal goal (time frame of goal); (4) learning versus performance goal; and (5) self-set versus assigned goal [82, 83, 84, 85].

First, goal difficulty has an effect on achievement. Locke and Latham [82] defined the *goal difficulty function* as the linear relationship between goal difficulty and performance. The goals of highest difficulty have the highest levels of effort and performance. Locke et al. [86] generally define the degree of difficulty of a goal as the probability that a goal can be reached. Goal difficulty has typically been set relative to the context of the goal. For example, LePine [87] set easy and difficult goals at approximately one standard deviation, respectively, below and above mean performance level. Mean performance level was captured per group across practice training trials. Earley [88] also collected preliminary data on subjects completing a particular task in order to assess goal difficulty. Results from this data found that 4% of subjects could complete 10 products in 15 minutes and that no subjects could complete more than 3 products in 5 minutes. With this data a *very difficult* goal was set as completing 10 products in 15 minutes for all participants in the study. In this dissertation study, goal difficulty is assumed to be relatively high for most participants given that I excluded students that are likely to be familiar with the search task topic.

Second, goal specificity has an effect on achievement. Results showed that *specific*, difficult goals consistently led to higher levels of performance [81]. Locke et al. [86] define the specificity of a goal in relation to the vagueness of a goal. While vague goals can be interpreted in various ways by different people, specific goals reduce variability of interpretation. This argument of specificity is in line with that of Hollenbeck and Klein [89]. They argue that "there are innumerable outcomes that could be consistent with a vague goal." [89, p. 214] Locke et al. [86, p. 272] offer a continuum of example goals that start vague and increase in specificity:

- **Most vague**: improve the performance of your division;
- **Less vague**: increase the profits of your division;
- **Less specific**: increase profits by 10% or more;
- **Most specific**: increase profits by exactly 15%.

This set starts with a goal that might have multiple interpretations (i.e., defining "performance" can take multiple forms) and ends with a goal that specifies goal components by limiting the number of allowed actions or outcomes (i.e., performance is specifically measured by profits that should increase by an exact percentage). Locke and Latham [81] argue that less specific goals or "do-your-best" goals have no external referent, being defined subjectively rather than from some clear objective resource. Locke and Latham also emphasize that goal specificity in and of itself (lacking difficulty) is not sufficient for high performance, but does reduce variation in performance and reduces ambiguity of the objective [86]. As an example, Latham and Seijts [90] used the following general goal versus specific and difficult goal:

- **General Goal**: Exerting high effort to make money typically results in high profits. Hence, it is important that you do your best to make as much money as possible.

- **Specific and Difficult Goal:** Exerting high effort to make money typically results in high profits. Hence, it is important that you commit to a specific difficult yet attainable goal to make money. In previous sessions, the average profit people earned was \$8.71. Your goal should be to make \$8.71 or more.

Third, the time frame of the goal has an effect on achievement. Goals can be either proximal (short-term) or distal (long-term). Proximal goals support achievement by increasing motivation and increasing self-efficacy through improved detection of errors [91]. Proximal goals have also been shown to be more helpful when overarching or distal goal is complex [92]. Distal goals should be broken down into proximal goals or subgoals: "Any distal or long-term goal can be segmented into several smaller and more immediate subgoals. These subgoals are intermediate steps to attaining the distal end goal and can be pursued in a sequential way." [83, p. 185] As an example, writing a literature review (distal goal) can be broken down into finding and categorizing relevant articles (proximal/subgoal), summarizing articles (proximal/subgoal), and developing an outline (proximal/subgoal). The completion of each subgoal can be used to track progress toward the larger, distal goal. Overall, results have shown that setting proximal goals in addition to a distal goal leads to higher achievement [90].

Fourth, learning goals lead to higher levels of achievement than performance goals. Elliot and Dweck [93] helped to define the categories of performance-oriented goals versus learning-oriented goals. Performance goals (sometimes referred to as *outcome* goals [84, 94]) are concerned with the judgment of a learner's ability to complete a particular task. In contrast, learning goals are concerned with developing a learner's ability over time in services of skill acquisition or task mastery. Further, "students with learning goals are interested in acquiring new skills and improving their knowledge, even if they make some mistakes. On the other hand, students with performance goals are usually interested in obtaining positive evaluations of ability and avoiding negative evaluations." [95, p. 72] Learning goals are established from a learner's intention to develop or improve a skill and are rooted in a desire to learn. Performance goals are established from a learner's intention to demonstrate competence to others and are rooted in a desire to obtain positive judgments.

Finally, the effect of self-set goals versus externally-set (assigned) goals has been investigated. Locke and Latham examined a series of 11 studies on the effects of self-set versus assigned goals on performance. Results found that "when goal difficulty is held constant, an assigned goal is as effective as one that is set participatively." [83, p. 10] However, they note that the logic or rationale behind assigned goals must be given for these results to hold true. When goals are self-set they are often selected by the probability of attainment determined by the learner. This probability is usually decided by ability or past experience. Those with high self-efficacy (high confidence in ability) tend to set higher goals than those with lower self-efficacy [82]. Individuals with high self-efficacy are also "...more committed to assigned goals, find and use better task strategies to attain the goals, and respond more positively to negative feedback than do people with low self-efficacy." [81, p.706] Locke and Latham also note two advantages to self-set goals in general. First, self-set goals tend to be more difficult which leads to higher performance [96]. Second, self-set goals can increase understanding of how to perform the task [97]. Additionally, Azevedo et al. assert that "allowing students to set learning goals can enhance their commitment to attaining them, which is necessary in order for goals to affect performance." [67, p. 6]

Locke and Latham outline four mechanisms accounting for the effect of goals on performance-

1. Goals direct attention and effort toward relevant activities and away from irrelevant activities (occurring cognitively and behaviorally)

2. Goals set at higher levels of difficulty lead to greater effort than goals set at a lower level (of difficulty)

3. Goals affect persistence, with more difficult goals prolonging effort

4. Goals affect action indirectly, instigating the "arousal, discovery, and/or use of task-relevant knowledge and strategies." [81, p.707]

Building on the last mechanism listed, there are several potential ways in which goals indirectly affect learners' actions. In response to goals, individuals reactively implement relevant skills and knowledge. If relevant skills are not automatic or readily available then individuals use skills from related contexts that may apply. If a task is completely novel individuals begin to develop new, relevant strategies for goal attainment. Self-efficacy impacts the effectiveness of these strategies. Self-efficacy refers to an individual's beliefs or confidence in one's ability to achieve a particular goal [98]. Individuals with high self-efficacy develop strategies that are more effective than those with low self-efficacy. Locke and Latham note that self-efficacy can be increased by ensuring adequate training, role modeling, and communication that exhibits confidence in an individual's ability to achieve the goal [99].

Through his exploration of the motivational influences of goals in academic settings, Zimmerman [100] further emphasized the findings of Locke and Latham. Zimmerman underscored that students with goals allocate attention toward task-relevant activities, exert greater effort, show greater persistence, have higher levels of attentiveness and self-satisfaction, and have lower levels of defensiveness.

Feedback plays a critical role in goal achievement. Feedback is the indicator of progress toward a goal. Without feedback, it may be difficult or impossible to engage in the metacognitive monitoring and control necessary to achieve a particular goal. If feedback indicates that goals are not being met individuals may changes strategies or increase their effort. "Feedback allows people to decide if more effort or a different strategy is needed to attain their goal. When performance feedback is withheld, goal setting is ineffective for increasing performance." [92, p. 7] Feedback can be internal or external. Internal feedback comes from metacognitive monitoring, comparing products or outcomes with standards, the criteria for optimal completion of a goal. If a discrepancy between these is identified, this internal feedback might prompt a learner to change tactics or strategies (metacognitive control). External feedback comes from outside the learner. This feedback might be delivered by teachers, teaching assistants, peers, or systems [101]. Butler and Winne note that prior work "generally confirms that learners are more effective when they attend to externally provided feedback." [102] In the next section, I explore how goal-setting impacts learning outcomes in general. I also investigate how particular goal characteristics impact learning.

2.3.1 Goal-Setting and Learning Outcomes

In the previous section, I focused on the effects that goals have on task performance. In this section, I focus on their effects on learning. Goals influence learners to engage in behaviors that benefit their learning—(1) goals direct attention *toward* relevant activities and *away* from non-relevant activities; (2) more difficult goals lead to greater effort exerted by the learner; (3) more difficult goals increase persistence, prolonging effort; and (4) goals instigate attentiveness, discovery, and use of task-relevant knowledge and strategies. One might question whether the relationship between goal-setting is causal versus correlational. If the relationship is causal, then difficult goals lead to greater effort. Conversely, if the relationship is correlational, then some other factor (e.g., prior knowledge) might influence learners to set difficult goals and exert greater effort. Prior research suggests that the relationship between goal-setting and effort is, in fact, causal. Citing their previous work [103, 104], Locke and Latham argue that once a goal is set, "effort is mobilized and expended in proportion to the difficulty level of the goal." [83, p. 6].

Sitzmann and Ely conducted a meta-analysis of SRL to identify gaps in the understanding of how learners self-regulate work-related knowledge and skills [33]. Sitzmann and Ely found that "goal level, persistence, effort, and self-efficacy were the self-regulation constructs with the strongest effects on learning." [33, p. 1] Across analyses of all studies, goals and self-efficacy were the only moderate-to-strong predictors of learning. Moeller et al. [105] conducted a longitudinal study of high school students investigating the effect of goal-setting on student achievement. "A correlational analysis of the goal-setting process and language proficiency scores reveals a statistically significant relationship between the goal-setting process and language achievement" [105, p. 153]. Across four years, each component of the goal-setting process (setting goals, creating actions plan for attainment, and reflection on attainment) correlated positively with each component of student assessment scores.

Studies have also investigated the effects of proximal (i.e., short-term) versus distal (i.e., long-term) goals on learning outcomes. In Morgan [106], students with proximal goals had better outcomes than students with distal goals. Additionally, students with proximal goals reported greater interest in the course. In a subsequent study, Morgan [107] found that students who engaged in more goal-setting and self-monitoring performed better on a final exam. Latham and Brown [94] investigated the effects of proximal and distal goals on MBA students across their first year of study. Results found that students who set specific, difficult, and proximal goals had higher GPAs and reported greater satisfaction with their performance.

Prior work has also demonstrated the impact of learning goals versus performance goals on learning outcomes. Schunk [108] investigated how goals affect motivation and achievement outcomes across two studies. One condition involved students with learning goals (i.e., aim is toward individual skill mastery) and the other involved students with performance goals (i.e., aim is toward proving ability). Results found that students with learning-oriented goals were more motivated and achieved better outcomes on a math problem-solving assessment. Additionally, students demonstrated greater *persistence* when they set learning-oriented goals and engaged in self-evaluation (i.e., judged their own ability to solve problems). McNeil and Alibali [109] investigated the effect of goal type on learning outcomes. Children in the study were either given learning goals, performance goals, or no goals toward problem-solving in math. Results found that children who were given learning goals were more likely to gain conceptual knowledge than those with performance goals. Further, McNeil and Alibali found that children who were given goals of any type were more likely to extend their knowledge beyond the taught mathematical procedure than the group of children not given goals. McNeil and Alibali argued that goals help problem solvers represent problems accurately.

2.3.2 Coding Goals

Prior work in goal-setting has coded self-set student goals along different dimensions to better understand the types of goals that improve achievement. McCardle et al. [39] developed four themes, collectively called TASC (i.e., time, actions, standards, and content), to assess micro-level goals that students set in a study session. The first three dimensions (i.e., time, actions, and standards) are rooted in the literature on ideal goal characteristics. First, time relates to both *specific* and *proximal* goal characteristics, shown to be optimal in prior work [83]. Second, actions relate to *specificity* as well. Goals with specific actions include cognitive process verbs, like identify, evaluate, or apply, that are engaged to complete the goal. This dimension stems from prior work [110], which has found that learning goals that involve specific actions (e.g., "discover n shortcuts" or "produce n schedules") improve outcomes during complex tasks. Participants in the study that completed these specific learning goals also reported higher rates of self-efficacy. Third, standards relate to the *specificity* of success criteria to measure goal achievement. This type of specificity provides a clear reference point for judging progress in achieving a goal. For example, prior work has found that a "complete n correct class schedules" goal has higher rates of performance and self-efficacy than a "do your best" goal [111]. The final dimension, content, was introduced by McCardle et al. [39]. They argue that content (or concepts) are the "foundation of effective learning goals because it focuses attention on the substance of learning rather than a sequence of tasks to complete." [39, p.2156] Further, they argue that specific content helps learners focus on relevant material and actions necessary for learning. Overall, goals with all four TASC criteria allow learners to—(1) monitor goal attainment; (2) know what and how they will learn; and (3) notice discrepancies between standards and current progress, making adjustments as needed.

2.4 Computer-Based Learning Environments

My research is centered on the role of goal-setting to support effective learning during search. To this end, I developed the Subgoal Manager as an auxiliary tool to encourage and facilitate goal setting and monitoring. Prior IR research has not investigated the effects of tools that specifically support goal-setting for learning during search. However, studies in the area of Computer-Based Learning Environments (CBLE) have evaluated stand-alone systems to support learning within specific domains. Many of these systems have included features explicitly designed to support effective goal-setting and other SRL processes [112, 113].

Azevedo et al. [114] developed a CBLE called MetaTutor. MetaTutor is an intelligent tutoring system and hypermedia-learning environment. MetaTutor is focused on helping learners to acquire conceptual knowledge, specifically complex biological content (e.g, circulatory, digestive, and nervous systems) [114, 115]. MetaTutor supports SRL processes in several ways. First, MetaTutor has four pedagogical agents (PAs) that guide and prompt students throughout the learning process. The first PA provides guidance and explanations of the MetaTutor environment from start to end. Additionally, it administers the pre- and post-test knowledge assessments and self-report measures. The second PA helps with and emphasizes planning, developing subgoals, and activating prior knowledge. The third PA prompts and supports monitoring processes. There are four main monitoring processes supported—(1) judgment of learning (i.e., encouraging learners to question whether they understood something they read); (2) feeling of knowing (i.e., encouraging learners to take inventory of their prior knowledge of a subject); (3) content evaluation (i.e., encouraging learners to evaluate the relevance of material based on their goals); and (4) progress monitoring (i.e., encouraging learners to assess whether a previously set goal has been met) [116]. The final PA prompts students to summarize content and gives feedback on the quality of summaries. Summary quality is judged by length and number of keywords.

2.5 Learning Assessments

Prior work in search-as-learning has used a *wide* variety of methods to measure learning. Some studies have measured learning by administering pre- and post-tests with predefined correct answers, including: (1) true-or-false [19, 24, 117, 16, 26, 118], (2) multiple-choice [19, 16, 119, 120, 22, 23, 121] and (3) short-answer tests [122, 25, 13, 29, 121, 21, 20]. Other studies have administered assessments with open-ended answers. Specifically, studies have measured learning by asking participants to: (1)

list relevant key phrases and facts [30, 18]; (2) create visual representations of a domain [17]; (3) enumerate arguments for and against a specific proposition [44]; and (4) summarize their knowledge of a topic [25, 29, 16, 123, 12, 31, 14, 124, 121, 11, 27]. To assess learning from open-ended responses, studies have adopted grading strategies that involve: (1) counting relevant concepts or facts [30, 11, 25, 29, 31, 18]; (2) counting relevant pro/con arguments [44]; and (3) counting statements that show evidence of generalization or critical thinking [25, 29, 12, 31, 27, 124]. Finally, studies have also considered self-reported perceptions of learning [25, 15, 17, 18, 120, 19] and behavioral measures that are assumed to provide evidence of learning [125].

Prior research has developed different coding schemes for open-ended responses. For example, Wilson & Wilson [126] developed a grading schemed for open-ended learning assessment responses. The grading scheme counts the number of statements (F-State) and facts (F-Fact) in a response. The scheme also considers the depth of responses assessing the quality of information across three dimensions, each inspired by a different cognitive process from the A&K taxonomy. For simplicity, in my dissertation, I focus on the percentage of (in)correct facts stated in the response.

Outside of search-as-learning, studies in the fields of psychology and education have used other types of assessments that may be worth considering in future search-as-learning research. In particular, there are two additional assessment types and scoring approaches that may be useful in search-as-learning studies: (1) task performance and (2) mental models.

Task performance assessments are designed to evaluate a learner's ability to carry out a complex task. Singley [127] investigated the effects of a specific system intervention added to a calculus tutoring system. To assess learning, task performance was measured from two perspectives: problem-solving speed and accuracy. While engaging with the tutor, participants solved calculus word problems. To measure problem-solving accuracy, each problem had a predefined *best* path with a specific sequence of moves. Problem-solving accuracy was measured by considering the number of unnecessary or illegal moves made by participants while solving problems. Participants with fewer unnecessary or illegal moves were considered to have better learning outcomes. Similarly, Koedinger and Anderson [128] explored the effects of a Cognitive Tutor for mathematical proofs. After a series of sessions with the Cognitive Tutor, learning was measured by asking participants to complete a series of proofs. Proofs were graded using a rubric adopted from Senk and Usiskin [129]. Each proof was given a binary score. Proofs were given a score of 1 if they had all the *key* steps correct (with or without minor errors in the details).

Mental models are subjective, cognitive representations of external reality [130]. Prior studies have used mental model assessments to better understand the level of conceptual knowledge a learner has acquired. Additionally, mental model assessments can illuminate gaps in an individual's understanding of a system or phenomenon. Chi et al. [131] used mental model assessment to measure learning during a tutoring session about the human body's circulatory system. Before and after the tutoring session, participants were asked to draw and explain the path of blood through the circulatory system on a sheet of paper with an outline of the human body. To analyze the drawings made by participants, Chi et al. developed seven different mental models. Six of these seven models had different degrees of errors. All seven models were ranked from the most naïve "No Loop" model to the most accurate and complete "Double Loop-2" model. Using this ranking of mental models, pre- and post-test drawings were analyzed in two distinct ways. First, the authors counted how many students had the correct "Double Loop-2" model before the tutoring session (0/11 students) and after the session (8/11 students). Second, the authors computed the average number of mental model shifts per individual student. For example, students who drew the most naïve "No Loop" model during the pre-test and the most complex "Double Loop-2" model during the post-test received a score of 6 (equal to the number of mental model "upshifts" from the pre-test to the post-test).

Finally, implementing multiple types of assessments in the same study can better capture depth and breadth of learning. McNeil and Alibali [109] is an example. McNeil and Alibali used a combination of assessments (i.e., short answer equivalence problems and open-ended questions) to measure: (1) conceptual learning, (2) procedural learning, and (3) transfer of learning. This combination of assessments enabled the researchers to measure breadth and depth of learning, as well as a learner's ability to transfer what was learned to solve a *new* type of problem (i.e., transfer of learning). In my dissertation study, I also implemented different assessment types. I administered a multiple-choice conceptual knowledge assessment and an open-ended summary of learning assessment.

CHAPTER 3

Research Questions

In my dissertation study, I explore the influences of goal-setting during search. In particular, I investigate how goal-setting affects learning during search and learning retention after search. Self-regulated learning (SRL) refers to an individual's ability to control their own learning. Effective SRL involves an active, reflective learner who is skilled at monitoring and controlling their own learning. Effective SRL has been shown to improve learning outcomes [32, 33, 2, 34, 35], therefore support of SRL should be an important area of focus for search-as-learning. As reflected in the Winne & Hadwin model of SRL [38], goal-setting is a critical phase of SRL. Self-regulated learners engage in goal-setting and continue to monitor their progress toward goal achievement throughout the learning process [71]. After many decades of research, the goal-setting literature has identified several important characteristics of goals that lead to higher levels of achievement. In particular, such goals are specific, difficult, proximal, and learning-oriented [84]. Additionally, goals that are self-set versus externally-set (assigned) tend to be more difficult (leading to higher performance) [96] and can increase understanding of how to perform the task [97]. Azevedo et al. [67] also argued that allowing individuals to set their own goals can increase motivation and commitment to attaining the goals. Goal-setting in general has several benefits that may positively impact learning during search—(1) goals direct attention and effort toward relevant activities; (2) goals lead to greater effort; (3) goals increase persistence; and (4) goals instigate use of task-relevant knowledge and strategies [81].

Capturing and measuring learning is difficult and complex. Through an extensive investigation of prior work in search-as-learning and the learning sciences [132], it is clear that a multi-dimensional approach is helpful to revealing what was learned and how well it was learned. For this reason, in my dissertation study I approach learning from several different dimensions. One dimension explores depth of learning immediately after search using an established two-tiered multiple-choice assessment (i.e., ODCA from Fisher et al. [1]). Another dimension explores the breadth of learning immediately after search and allows for more flexibility in expressing what was learned with an open-ended assessment. Yet another dimension explores learning retention by measuring learning with both assessments one week after the search session. Finally, another dimension investigates rates of SRL through interactions and think-aloud data. This study explores the impact of a goal-setting condition on each of these learning dimensions. Additionally, this study explores the impact of a goal-setting condition on searcher perceptions and search behaviors.

In this chapter, research questions are provided with hypotheses and rationale. Each of the research questions explores the influence of subgoal conditions on various factors. To review, there are two subgoal conditions in my dissertation study:

- SUBGOALS: Participants were provided with The Subgoal Manager containing blank fields to develop subgoals, Figure 1.1. Subgoals could be added and deleted throughout the search session. Participants were informed about qualities of good subgoals (i.e., subgoals that are more likely to be achieved). Participants were asked to develop at least 3 subgoals before beginning the search session.

- NOSUBGOALS: Participants were provided with the Text Editor. Participants were not informed of subgoal qualities nor asked to develop subgoals.

3.1 Research Question 1 (RQ1)

RQ1: How does the subgoal condition influence learning during search?

RQ1 investigates the impact of different subgoal conditions (i.e., SUBGOALS and NOSUBGOALS) on learning. Prior work in search-as-learning and the learning sciences has emphasized the usefulness of multiple assessment types to capture learning [132]. Measurement of learning was recorded with two different assessments at two different points in time. One learning assessment is open-ended and captures the breadth of what was learned. The second learning assessment (i.e., ODCA from Fisher et al. [1]) is more narrow, capturing the depth of learning of specific concepts. The assessments help to unpack *what* was learned. Measuring learning at different points in time explores how well participants *retained* what was learned. Within **RQ1**, I investigate the two following research sub-questions that explore the impact of subgoal conditions:

RQ1-A: On post-task learning scores?

RQ1-B: On delayed learning scores (i.e., retention)?

I examine the following hypothesis using scores from the immediate post-task and retention learning assessments:

H1: Participants will have higher learning outcomes in the SUBGOALS condition.

Rationale: The above hypothesis presumes higher scores on the multiple-choice ODCA assessment and the open-ended learning assessment in both immediate post-task scores and retention scores in the SUBGOALS condition. Hypothesis H1 is motivated by two trends observed in my preliminary study (Appendix A). Namely, in the SELFSETSUBGOALS condition, participants had higher scores on a post-task assessment (i.e., the ODCA [1]) and reported greater levels of SRL support from having access to the Subgoal Manager. Additionally, prior work has demonstrated that learners who set goals toward their learning task are likely to have better learning outcomes than those who do not set goals [33, 105]. The above hypothesis is also partially based on higher rates of perceived support of SRL in the SELFSETSUBGOALS condition in the preliminary study. Better support of SRL may support better long-term learning. During effective SRL, learners activate prior knowledge and integrate what they are learning with existing long-term knowledge. There is also evidence that differences between groups may increase in retention assessments when compared with immediate post-task assessments. Prior work has shown limited differences between groups in immediate post-task learning assessments with greater differences between groups in retention learning assessments [133].

3.2 Research Question 2 (RQ2)

RQ2: How does the subgoal condition influence searcher perceptions? I investigate perceptions of interest, knowledge increase, satisfaction, difficulty, SRL support, and engagement.

RQ2 investigates the effect of subgoal condition on searcher perceptions.

H2: Participants will report higher levels of SRL support in the SUBGOALS condition.

Rationale: Hypothesis H2 is rooted in prior work and the preliminary study. Prior work in SRL explains goal-setting as an important component of effective SRL. Additionally, the preliminary study showed that participants in the SelfSetSubgoals condition reported higher rates of SRL support in planning and evaluating subgoal progress than in the NoSubgoals condition. Uniquely, participants in the SelfSetSubgoals condition reported higher levels of adapting strategies or approaches.

The effects of the subgoal condition on perceptions of interest, knowledge increase, satisfaction, difficulty, and engagement are exploratory. On one hand, prior work has found that allowing individuals to set their own goals can increase their motivation and commitment to attaining the goals [67]. This may cause participants in the Subgoals condition to experience higher interest, perceived knowledge increase, satisfaction, and engagement with lower levels of difficulty. On the other hand, the use of a novel tool, the Subgoal Manager, may introduce a heavier cognitive load than the simple Text Editor with which participants are familiar. It is possible this heavier cognitive load may result in higher levels of difficulty and lower levels of satisfaction, interest, perceived knowledge increase, and engagement. Concerning knowledge increase in particular, greater motivation may lead to greater perceptions of learning. Conversely, greater motivation may influence participants to increase personal standards, leading to lower perceptions of learning even if they *objectively* learned more (**RQ1**).

3.3 Research Question 3 (RQ3)

RQ3: How does the subgoal condition influence search behavior (e.g., queries, clicks, etc.) during search?

RQ3 investigates the effect of subgoal condition on search behaviors. A central goal of search-as-learning is to better understand which search behaviors may indicate that learning is occurring during search. For this reason, search-as-learning has investigated search behavior during learning-oriented search tasks. Prior work has investigated the correlation between certain behavioral measures and learning outcomes [24, 25, 26, 27, 28, 29, 30, 31]. However, behavioral measures are sometimes difficult to interpret. For example, Collins-Thompson et al. [25] found participants with better learning outcomes spent more time per document clicked on the SERP. Bhattacharya & Gwizdka [30] conducted a deeper analysis of participants' reading behaviors. Results found that participants with higher learning outcomes had shorter reading sequences and fewer eye regressions. Shorter reading sequences might suggest that these participants sought specific content based on their goals. Fewer eye regressions suggest that these participants had less difficulty understanding the content.

Combined, these results suggest that behavioral measures can be difficult to interpret in isolation. There are many underlying factors that influence whether or not a particular search behavior indicates learning. Many of these factors are dependent on the individual learner (e.g., experts issued more queries than novices [134]), the learning-oriented task (e.g., more sources were visited in more complex goals [135]), and the search system (e.g., more queries were issued on laptops than smartphones [44]). Because there are so many factors that can influence search behaviors during learning-oriented search tasks, **RQ3** is purely exploratory and does not have a fixed hypothesis. Generally, this research question investigates how subgoal conditions (i.e., SUBGOALS or NOSUBGOALS) influence how people search while learning when controlling the type of learning-oriented search task.

3.4 Research Question 4 (RQ4)

RQ4: How does the subgoal condition influence SRL processes? I investigate *Planning* (e.g., subgoals, prior knowledge activation, and recycle goal in working memory), *Strategy Use* (e.g., taking notes, reading notes, and comparing & contrasting), and *Monitoring* (e.g., monitoring progress toward subgoals, content evaluation, and judgment of learning).

RQ4 investigates the effect of subgoal conditions on SRL processes. Goal-setting is a critical component of effective SRL. Goals prompt learners to—(1) consider their task understanding; (2) direct attention toward planning and strategy choice for goal achievement; and (3) provide standards for monitoring and evaluating progress [39]. During SRL, goals act as long and short term sources of feedback in metacognitive monitoring and control. Prior work in goal-setting has shown that goals affect action, instigating the use of task-relevant knowledge and strategies [81]. Locke and Latham also note that self-set goals in particular can increase understanding of how to perform the task [97]. As discussed in Chapter 2, SRL self-report questionnaire data is quite coarse and often unreliable. To better understand the frequency and types of dynamic SRL processes that occur during the search session, think-aloud and interaction data was coded as macro-SRL and micro-SRL processes (definitions of specific macro-SRL and micro-SRL processes are provided in Tables 4.11, 4.13, 4.15, and 4.17).

> **H4:** Participants will engage more frequently in SRL processes in the SUBGOALS condition.

Rationale: Hypothesis H4 is rooted in prior work from SRL and goal-setting theory. Participants in the SUBGOALS condition were asked to develop their own subgoals toward the overall learning objective. The process of subgoal development itself should have several benefits that include higher rates of *Planning* micro-SRL processes (e.g., making subgoals and prior knowledge activation to develop relevant and challenging subgoals). Having developed their own subgoals, participants in the SUBGOALS condition should also engage in higher levels of metacognitive *Monitoring* micro-SRL processes as the goals will act as sources of feedback or standards for measuring progress. Additionally, participants in the SUBGOALS condition should be more motivated to achieve subgoals, directing more attention toward strategy selection, potentially resulting in greater levels of particular micro-SRL processes of *Strategy Use* that are found to be useful in completing this particular overall search task.

CHAPTER 4

Methods

To investigate the four research questions described in Chapter 3, I conducted a lab study via Zoom with 40 participants. All participants were undergraduate students at the University of North Carolina at Chapel Hill. Participants were assigned to one of two subgoal conditions: SUBGOALS or NOSUBGOALS. In both conditions, participants completed a single learning-oriented search task and thought aloud while searching. In the SUBGOALS condition, participants used the Subgoal Manager to complete subgoals toward the overall learning objective. In the NOSUBGOALS condition, participants were provided with the Text Editor to take notes, but were not informed about subgoals. Participants also completed pre-task and post-task learning assessments and returned one week later to complete the learning retention assessments.

This section provides an overview of the methodology of the study. First, I provide a detailed description of the study protocol. Next, I discuss the sample size, population, and recruitment of the study participants. Then, I provide a description of the instructional videos used in each subgoal condition. Next, I provide the learning-oriented search task (and example task used for participants to practice thinking aloud). Next, I describe the search system used in the study. Then, I provide descriptions and items in each of the three questionnaires used in the study—(1) demographics questionnaire; (2) pre-task questionnaire; and (3) post-task questionnaire. Next, I discuss the multiple-choice and open-ended learning assessments. Then, I describe the qualitative analysis of the open-ended assessments. In the subsequent section, I discuss the search behaviors, think-aloud comments, and screen recordings that were collected. Next, I describe the qualitative analysis in which SRL processes were identified from think-aloud comments and screen recordings. Finally, I describe the statistical analysis that was conducted.

4.1 Study Protocol

An overview of the complete study protocol workflow is shown in Figure 4.1. This study took place over two sessions (conducted one week apart). The first session began by providing the participant with a brief overview video of the study, followed by a consent form that was signed by each participant before continuing the study. Next, each participant completed a brief demographics questionnaire, shown in Table 4.1. After the demographics questionnaire, each participant engaged in a five-minute think-aloud practice task. The moderator prompted the participant by saying, "Please keep talking," as necessary throughout the practice task. Then, participants completed a 18-item (or 9 pair) 2-tiered multiple-choice learning assessment (Section 4.8). Next, participants read the learning-oriented task description and completed a pre-task questionnaire with items that asked about their perception of the task, (Table 4.2).

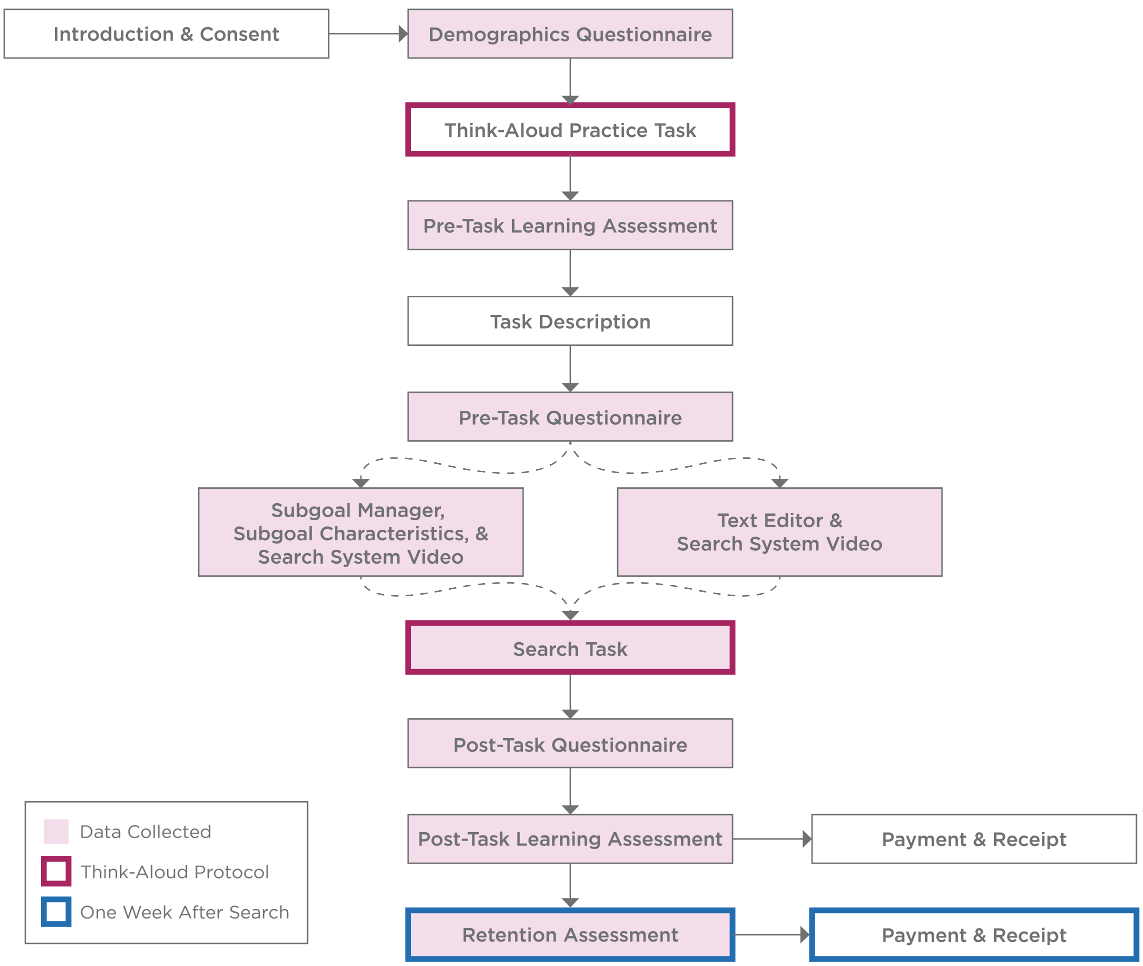

Figure 4.1 Study Protocol

Next, each participant viewed one of two videos based on the subgoal condition (i.e., Subgoals or NoSubgoals) (Section 4.4). Next, each participant completed the learning-oriented search task using their respective auxiliary search tool (i.e., Subgoal Manager or Text Editor). Participants were given a maximum of 40 minutes to complete the task. After completing the search session, participants completed a post-task questionnaire with items about their perceptions of their search experience (Tables 4.4- 4.10). After the post-task questionnaire, participants were given up to 10 minutes to complete an open-ended learning assessment question (Section 4.5) followed by the same pre-task multiple-choice assessment.

To navigate through each step of the study, participants used the "Study Workflow" interface (Figure 4.2). The interface included a button for each step. The current active step was clickable and highlighted in blue. All other steps were not clickable.

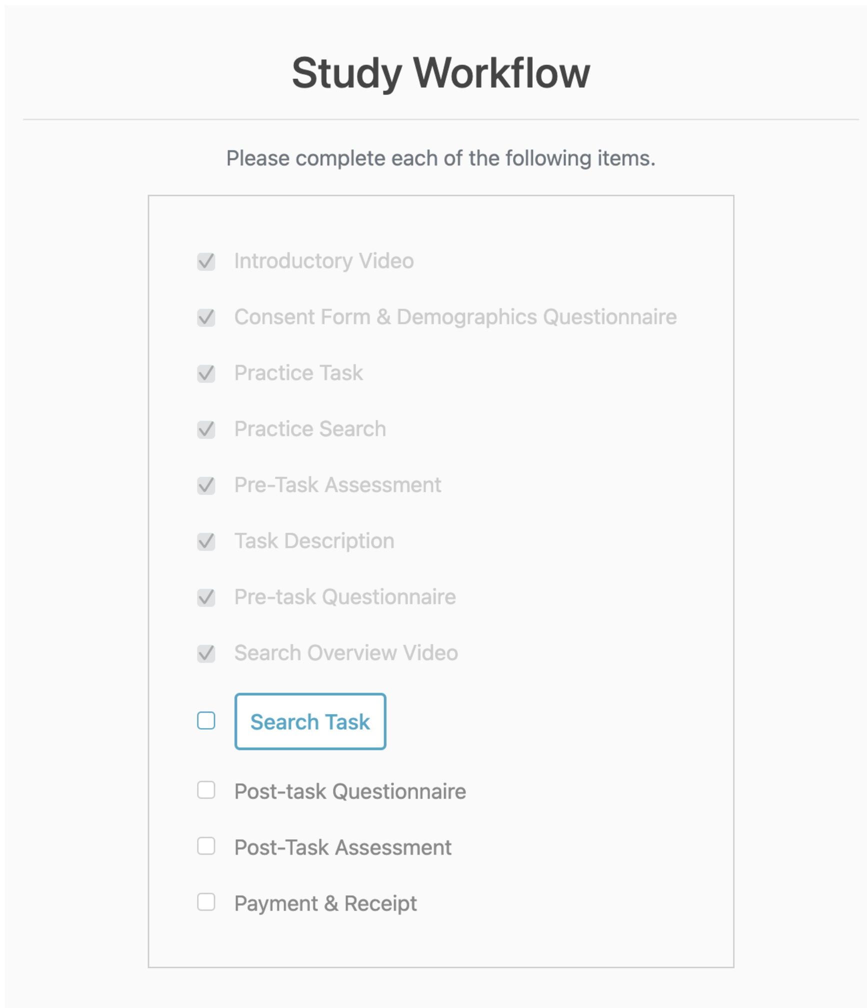

Figure 4.2 Study Workflow interface that participants used to navigate the study ("Search Task" is highlighted in blue to indicate the current active phase, all other phases are not clickable).

Participants were paid US$30.00 for the initial session of about 75 minutes. One week later, participants completed a retention assessment consisting of the same open-ended question and

18-item multiple-choice assessment. Participants were paid US$10.00 for this follow-up session for a total of US$40.00 across both sessions.

4.2 Power Analysis

For this study, a post hoc power analysis indicated that a sample size of 40 participants (20 participants per subgoal condition) provided Shieh Power of 0.8 for between-subjects effects of learning assessment score using Wilcoxon Rank Sum Test assuming $\alpha = .05$, exponential distribution, and an effect size $f = 0.72$. Power analysis was conducted for Wilcoxon Rank Sum test using the R package wmwpow following methods and recommendations of Mollan et al. [136].

4.3 Participant Recruitment

I recruited 40 University of North Carolina at Chapel Hill (UNC) undergraduate students (28 identified as female and 12 identified as male) for the study. Participant ages ranged from 18 to 22 years, with an average age of 19.93 years (SD = 1.07). I focused on the undergraduate student group because it had the largest population among UNC students, faculty, and staff. I focused on a single group in order to reduce the likelihood of variance between participant groups. Also, I selected the most populous group at UNC to make recruitment most efficient. Participants were representative of first through fourth year students with 7 first-year students, 6 second-year students, 12 third-year students, and 15 fourth-year students.

In addition to the stipulation of UNC undergraduate students, participants were required to be fluent in written and spoken English and included only non-biology and non-chemistry majors. The central concepts of the search task were diffusion and osmosis. These concepts are included in undergraduate biology courses required for all biology and chemistry students. The point of the learning-oriented search task is to present a challenging topic that is somewhat unfamiliar, therefore undergraduate students majoring in biology and chemistry were not included as participants in this study. Participants in the study came from various academic departments. In total, participants were representative of 20 distinct majors. Participants also had a variety of experience in biology: 24 had taken an undergraduate-level course in biology; 14 had taken a high school-level course in biology; 1 had taken a graduate-level course; and 1 had not previously taken a course in biology.

Each participant completed a single learning-oriented search task (discussed in Section 4.5) in one of the two subgoal conditions, either Subgoals or NoSubgoals. They were asked to attend an initial 75 minute session that included the search session, questionnaires, and pre- and post-task learning assessments, along with a second follow-up session one week later to complete a learning retention assessment that took about 15 minutes.

4.4 Instructional Videos

Based on the experimental condition, participants were presented with one of two instructional videos before the search task. In the SUBGOALS condition, the instructional video contained three main components. First, the video provided a brief tutorial of the Subgoal Manager used to take notes and save information while searching on each subgoal. Next, the video explained that there are ideal subgoal characteristics that make subgoals more achievable. Additionally, the video explained four ideal subgoal characteristics, specific: time; action; standard; and content [39]. In the video, I provided definitions of each characteristic along with the following example subgoal that contained all 4 characteristics: "Spend 10 minutes [*specific time*] identifying [*specific action*] three paintings [*specific standard*] that demonstrate the main characteristics of surrealism [*specific content*]." Finally, the video informed participants that they would be asked to develop at least three subgoals and that they should incorporate these ideal subgoal characteristics in their subgoals.

In the NOSUBGOALS condition, the video had one main component. The video provided a brief overview of the Text Editor used to take notes and save information while searching.

The NOSUBGOALS video was about two minutes long and the SUBGOALS video was about six minutes long, meaning participants in the SUBGOALS condition spent about four minutes longer during the instructional video phase.

4.5 Learning-Oriented Search Task

Before the main search task, each participant completed an example task to practice the think-aloud technique. The think-aloud practice task was also a learning-oriented search task of a similar structure:

> *Scenario*: After a recent trip to the Museum of Modern Art you become fascinated with an exhibit featuring art movements of the 20th century. In particular, you become interested in surrealism and dadaism.
>
> *Task*: Gather information and **learn everything you can about the concepts of *surrealism* and *dadaism***.

The goal of my dissertation study is to investigate the impact of subgoals on learning during a search session. The focus of my research is the impact of subgoals on a complex learning objective. In a previous study [137], I investigated the impact of a variety of complex learning objectives on participant perceptions and learning pathways traversed. In the study, I found that conceptual learning objectives were perceived as the most difficult and required more search activity. I chose a difficult conceptual learning task because I expected participants to benefit from goal-setting support to achieve success. Additionally, prior work on goal-setting has found that *difficult* goals increase performance. In this respect, I also expected participants to set difficult subgoals during a difficult conceptual learning task.

The conceptual learning-oriented search task is as follows:

> *Scenario*: One of your family members is a high school senior who is about to take an important biology exam. Your family member has told you that she is struggling to understand the concepts of diffusion and osmosis and has asked for your help.
>
> *Task*: Gather information and **learn everything you can about the concepts of *diffusion* and *osmosis***. After searching for and gathering information, you will be asked to answer some questions about both diffusion and osmosis.

4.6 Search System

The search system was implemented using the Bing Web Search API. Given a query, the system returned results from four different verticals in different tabs: web, images, news, and video. Each vertical tab displayed the top-50 results. As shown in Figure 4.3, the vertical tabs were displayed along the top. The web vertical was the default vertical. The web, videos, and news verticals included pagination controls at the bottom and displayed 10 results per page. The Bing API was configured to return results for the US-EN market and had safe-mode turned on to filter inappropriate results. At the top of the search interface were three buttons. The first button opened the Subgoal Manager in the Subgoals condition and the Text Editor in the NoSubgoals condition. The next button opened the task description in a new tab and could be used when participants wanted to re-read the task description. The last button was used when participants were done with the task. This button excited the search interface and routed participants to the post-task questionnaire.

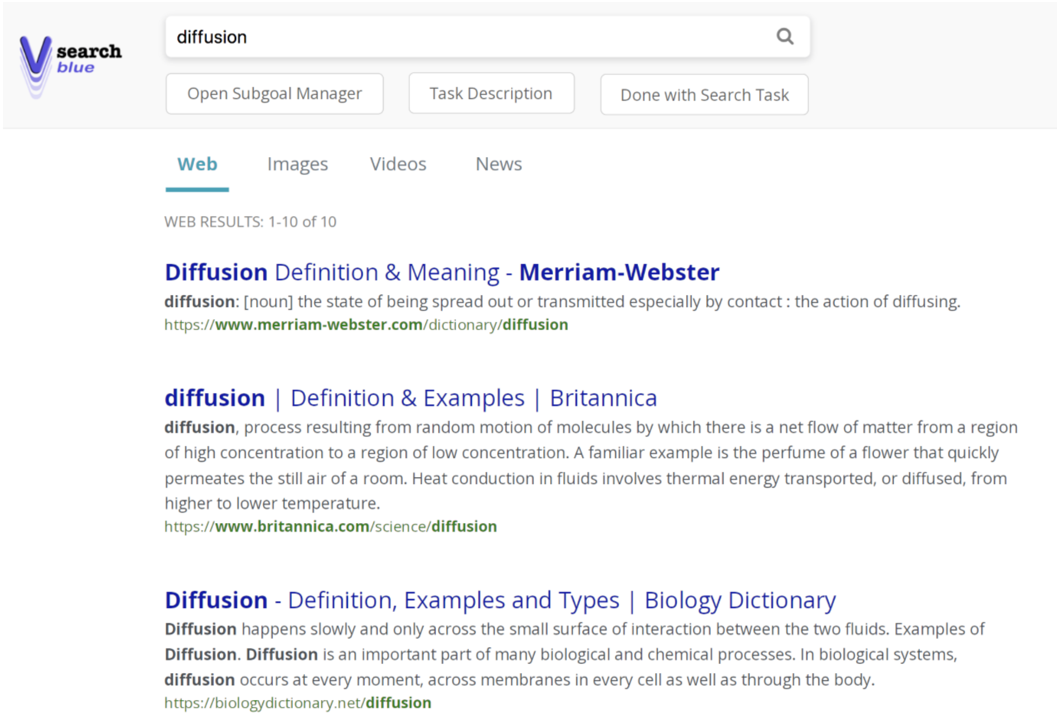

Figure 4.3 Search interface in SUBGOALS condition (in NOSUBGOALS condition the “Open Subgoal Manager” button is changed to “Open Text Editor”).

4.7 Questionnaires

The study included three questionnaires—(1) a demographics questionnaire (Table 4.1); (2) a pre-task questionnaire (Table 4.2); and (3) a post-task questionnaire with four parts (Tables 4.4- 4.10). Data from the pre-task questionnaire was collected to verify there were no significant differences between groups (i.e., reported prior knowledge, a priori determinability). To address **RQ2**, the post-task questionnaire asked about participant perceptions of the task and search experience. The following section provides an overview of each questionnaire and details all questionnaire items.

4.7.1 Demographics Questionnaire

After completing a consent form, participants completed a brief demographics questionnaire. Table 4.1 shows the demographics questionnaire items. In total, there are four questionnaire items—one fill-in-the-blank age question; one checkbox and fill-in-the-blank gender identity question (developed using recommendations from Fraser [138]); and two multiple-choice education questions.

Table 4.1 Demographics questionnaire

Theme	Items
Age	What is your age? [*fill in the blank*]
Gender Identity	What is your gender identity? Select all that apply. • Female • Male • Trans female/trans woman • Trans male/trans man • Non-binary • Not listed (please state) [*fill in the blank*]
Education	What is the highest degree or level of schooling you have completed? • No schooling completed • Elementary school to 8th grade • Some high school, no diploma • High school graduate, diploma or equivalent (for example: GED) • Some college credit, no degree • Trade/technical/vocational training • Associate degree • Bachelor's degree • Master's degree • Professional degree • Doctorate degree What is the highest level biology course you have completed? • No biology courses completed • Elementary school to 8th grade level biology course • High school level biology course • Undergraduate level biology course • Graduate level biology course

Table 4.2 Pre-task questionnaire (all items are Likert scale from 1 [strongly disagree] to 7 [strongly agree]).

Theme	Items
Interest	I am interested in this topic.
Prior Knowledge	I know the scientific definition of diffusion. I know how diffusion works in great detail. I know the scientific definition of osmosis. I know how osmosis works in great detail. I know a lot about the overall task topic.
Difficulty	I think it will be difficult to decide when I have enough information to complete this task. I think it will be difficult to complete this task. I think it will be difficult to search for information to complete this task. I think it will be difficult to integrate the information I find to complete this task.
A Priori Determinability	Right now, I know what my approach to this task will look like. Right now, I know what information might be useful to complete this task. Right now, I know what goals I need to accomplish in order to address the task. Right now, I understand what I need to learn. Right now, I know where to begin my search. Right now, I understand what this task is asking me to do. Right now, I understand the terminology in the task description.

4.7.2 Pre-Task Questionnaire

After reading the learning-oriented search task, participants completed a pre-task questionnaire about their perceptions of the task. Table 4.2 shows the pre-task questionnaire items. All items are Likert scale from 1 (strongly disagree) to 7 (strongly agree). In total, there are 17 questionnaire items—1 interest question; 5 prior knowledge questions; 4 difficulty questions; and 7 a priori determinability questions. The 5 prior knowledge questions include 2 questions about diffusion, 2 questions about osmosis, and 1 general question about the overall task topic. The interest, difficult, and a priori determinability questions have been used in prior work [139, 137].

4.7.3 Post-Task Questionnaires

After completing the learning-oriented search task, participants completed a post-task questionnaire that covered six themes. The initial four sections measure interest, self-assessment of learning, satisfaction, and difficulty and are detailed in Table 4.4. While the interest, satisfaction, and difficulty items have been used in a previous study [137], the self-assessment of learning items were new to

Table 4.4 Post-task questionnaire (all items are Likert scale from 1 [strongly disagree] to 7 [strongly agree]).

Theme	Items
Interest	My interest in this topic has increased.
Learning Self-Assessment	After searching, I know more about the definition of diffusion.
	After searching, I know more about how diffusion works.
	After searching, I know more about the definition of osmosis.
	After searching, I know more about how osmosis works.
	After searching, I know more about the task topic.
Satisfaction	I am satisfied with the amount of information I found to complete this task.
	I am satisfied with the strategy I took to find information for this task.
	I am satisfied with the quality of information I found to complete this task.
	I am satisfied with the amount of time I spent on this task.
Difficulty	It was difficult to decide when I had enough information to complete this task.
	It was difficult to complete this task.
	It was difficult to search for information to complete this task.
	It was difficult to integrate the information I found to complete this task.

Table 4.6 Post-task questionnaire subgoal adherence and achievement items (all items are Likert scale from 1 [strongly disagree] to 7 [strongly agree]) presented in the Subgoals condition.

Theme	Items
	During the search task:
Subgoal Adherence & Achievement	I stuck to the set of subgoals.
	I searched outside of the set of subgoals.
	I fully completed the set of subgoals.

this study. These items mirror the prior knowledge items from the pre-task questionnaire, shown in Table 4.2.

The next section of the post-task questionnaire was only presented to participants in the Subgoals condition. As shown in Table 4.6, this section asked three questions about the extent to which participants adhered and achieved their subgoals.

Next, participants completed one open-ended question about their strategies and approaches to the learning-oriented search task:

> Describe the strategy and/or approach you used to complete the search task (2-5 sentences):

Table 4.8 Post-Task questionnaire self-regulated learning items designed to capture core components of the Winne & Hadwin [2] model of SRL (all items are Likert scale from 1 [strongly disagree] to 7 [strongly agree]).

Theme	**Items**
	During the search task, I was able to:
Planning	Set goals for what information I needed to find. Develop a plan for exploring the topic. Decide where to begin my search. Decide how to approach the task.
Strategy Use	Make connections between the information I found and my own prior knowledge. Consider connections between ideas. Consider how parts of the topic relate to each other. Structure my thoughts.
Monitoring	Question whether I would be able to recall the information I encountered. Question whether I understood the information I encountered. Question whether I understood the information I needed to know to complete the task. Question whether I knew what was important to complete the task.
Evaluating Progress	Decide when to transition from one goal to the next. Decide when a goal was successfully accomplished. Keep track of my progress toward goals. Organize the information that I found. Understand the structure of the topic related to the task.
Adapting	Change or revise my understanding of the task itself. Change or revise my level of confidence during the task. Change or revise my approach to the task. Change or revise how much effort I devoted to a specific goal. Change or revise my feelings about my approach to the task.

& Hadwin). The next four items capture monitoring learning (Monitoring of Winne & Hadwin). The next five items capture evaluating progress (Monitoring and Control of Winne & Hadwin). The final five items capture adapting behaviors (Phase 4 Adaptations of Winne & Hadwin).

Finally, participants were presented with questionnaire items that targeted search engagement. All items shown in Table 4.10 were adapted from O'Brien [3]. The first three items target perceived usability, how taxing, confusing, and frustrating the search experience was perceived to be. The next three items target reward, how worthwhile, interesting, and rewarding the search experience was perceived to be. The final three items target focused attention, how absorbed the participants were in the search experience. Typically, there are four dimensions of search engagement. However, because the study focused on the strategy of goal-setting, rather than a particular auxiliary search

Table 4.10 Post-task questionnaire engagement items adapted from O'Brien et al. [3] (all items are Likert scale from 1 [strongly disagree] to 7 [strongly agree]).

Theme	Items
Perceived Usability	I found the search experience confusing.
	The search experience was taxing.
	I felt frustrated while searching.
Reward	The search experience was worthwhile.
	I felt interested in the search task.
	The search experience was rewarding.
Focused Attention	The time I spent searching just slipped away.
	I lost myself in the search experience.
	I was absorbed in the search task.

tool, the tool-specific dimension of aesthetic appeal was not included.

4.7.4 Cronbach's Alpha

A Cronbach's α analysis was conducted to determine if visually grouped questions that were designed to capture a particular construct had high internal consistency. A minimum threshold of $\alpha = 0.7$ was set for items to be considered internally consistent.[1] Groups of items that had an $\alpha \geq 0.7$ were combined into a single variable by calculating the mean across the set of items.

I calculated Cronbach's α across 3 groups of items from the pre-task questionnaire (prior knowledge, expected difficulty, and a priori determinability) and 11 groups of items from the post-task questionnaire (knowledge increase, satisfaction, perceived difficulty, SRL *Planning*, SRL *Strategy Use*, SRL *Monitoring*, SRL *Evaluating Progress*, SRL *Adapting*, perceived usability, reward, and focused attention).

In two cases, SRL perceptions of *Strategy Use* and *Monitoring*, groups of items on the post-task questionnaire had an $\alpha < 0.7$. In these cases, one item was dropped to increase α above 0.7. Within the set of items relating to *Strategy Use*, the item "Structure my thoughts." was dropped. Within the set of items relating to *Monitoring*, the item "Question whether I would be able to recall the information I encountered." was dropped. In one case, perceived difficulty, α was less than 0.7 regardless of which items were dropped. In this case, each item of perceived difficulty was analyzed separately.

[1] $\alpha = 0.7$ is a typical acceptable level for items to be considered internally consistent [140]. While high levels of Cronbach's α do not imply unidimensionality, low levels of Cronbach's α indicate a high level of error variance and are more likely to be multidimensional [141].

4.8 Learning Assessments

To address **RQ1**, participants were asked to complete learning assessments at three different stages: before the search session (pre-task learning assessment), after the search session (post-task learning assessment), and again one week after the search session (learning retention assessment). Capturing learning at three points of time was intended to make clear what participants knew before searching, immediately after searching, and what they retained after searching.

The pre-task, post-task, and retention learning assessments consisted of a two-tiered multiple-choice assessment developed by Fisher et al. [1], called the Osmosis and Diffusion Conceptual Assessment (ODCA). The ODCA is based on Odom & Barrow's Diffusion and Osmosis Diagnostic Test (DODT) [142]. Odom & Barrow found that biology majors exhibited low performance on the DODT, with non-biology majors and high school students having even lower scores. These findings demonstrated the difficulties for students to master the concepts of osmosis and diffusion. Similarly, Fisher et al. report that only a fraction of college students demonstrated full understanding of the mechanisms of diffusion and osmosis, measured by performance on the ODCA. Further, biology majors *did not* perform better than non-majors on the ODCA.

I have selected the ODCA for two main reasons. First, the ODCA targets concepts that have known misconceptions and are challenging even for those students specializing in the topic. Second, the ODCA is a valid and reliable tool for assessing understanding of osmosis and diffusion [1]. Fisher et al. conducted multiple iterations of face validation with expert biology faculty, modifying the language of response choices throughout. Additionally, 16 undergraduate students participated in semi-structured interviews where they read ODCA items aloud and provided explanations for why they would select or reject each response. The students were also asked to suggest alternative responses if they were not satisfied with those offered. To establish construct validity and reliability, Fisher et al., "calculated the standard difficulty index (p: proportion of students who answered a test item correctly) and discrimination index (d: index that refers to how well an item differentiates between high and low scorers), and evaluated the instrument's reliability with Cronbach's alpha [...] using PASW Statistics" [1, p. 420]

The ODCA includes 18 individual questions that are organized as 9 pairs. Each pair contains a knowledge question and a reasoning question. A knowledge question aims at the content knowledge or the "What?". The reasoning question aims at the explanation or the "Why?" associated with the content knowledge question. As an example, Figure 4.4 shows the first two questions on the ODCA. The first (knowledge) question tests the participant's understanding of cell membranes. The second (reasoning) question tests the participant's justification for selecting semipermeable or permeable as the answer.

Figure 4.4 Example learning assessment question pair from the ODCA by Fisher et al. [1].

1. All cell membranes are:
   A. semipermeable
   B. permeable
2. The reason for my answer is because cell membranes:
   A. allow free movement of materials into or out of the cell.
   B. allow some substances to enter the cell, while they prevent all substances from leaving.
   C. allow only beneficial materials to enter the cell.
   D. allow some substances to pass through, but not others.

In addition to the ODCA, the post-task and retention learning assessments also contained a single open-ended question designed to capture each participant's knowledge of the concepts central to the learning-oriented search task. Shown in Figure 4.5, the open-ended question asked participants to write everything they learned about the concepts of osmosis and diffusion. Participants were also encouraged to include as much relevant information as possible even if they were not fully confident. This was added to increase the likelihood of capturing *everything* participants learned or retained from the search session.

Figure 4.5 Open-ended learning assessment question used in post-task and retention learning assessment phases.

In 5-8 sentences, write everything you learned about diffusion and osmosis during the search session. Please include as much relevant information and detail as possible even if you are not fully confident.

To summarize, the post-task learning assessment and learning retention assessment were identical. The assessments included both the open-ended question, shown in Figure 4.5 and the ODCA multiple-choice assessment.

4.9 Normalized Gain

The normalized gain was calculated for both post-task and retention outcomes from the ODCA assessment. Normalized gain is calculated as:

$$NormalizedGain = \frac{(PostTaskScore - PreTaskScore)}{(1 - PreTaskScore)} \tag{4.1}$$

where $PostTaskScore$ is equal to the number of questions correct on the post-task ODCA divided by the total number of questions and $PreTaskScore$ is equal to the number of questions correct on the pre-task ODCA divded by the total number of questions. The same normalized gain was calculated for the retention assessment, where $RetentionScore$ was substituted for $PostTaskScore$ with $RetentionScore$ equal to the number of questions correct on the retention ODCA divided by the total number of questions.

This type of normalization is common among education research [143] and search-as-learning studies [24, 26, 144]. The normalized gain accounts for what a learner already knew on the pre-task assessment and, essentially, answers the following question: "Of the percentage a learner could gain, what percentage did they gain?"

4.10 Qualitative Coding of Open-Ended Learning Assessments

Each participant completed two open-ended assessments in which they were asked to recall everything they learned during the search session. One assessment was completed immediately after the search task and another was completed one week later. This resulted in 80 open-ended responses (i.e., 40 participants x 2 responses per participant).

Ultimately, our goal was to analyze each open-ended response as a set of correct and incorrect statements (i.e., percentage of correct statements). To this end, we had to: (1) identify all the distinct statements included in a response and (2) know the veracity of each statement. Our qualitative analysis of open-ended responses involved several steps, described in detail below. Each step was validated by: (1) developing a coding guide; (2) having two researchers (R1 and R2) perform tasks independently; (3) measuring agreement; and (4) repeating the process until agreement was sufficiently high.

4.10.1 From Sentences to Statements

Analyzing open-ended responses as a set of correct and incorrect statements required identifying the set of all unique statements in each response. This process involved: (1) automatically splitting each response into sentences using a period as the delimiter; (2) manually identifying all the logical atomic statements (referred to as simply "statements") in each sentence; and (3) taking the union of all distinct statements in the response.

Sentences could contain one, two, or several logical atomic statements. Consider the following sentence:

> "Diffusion and osmosis are forms of passive transport."

This sentence contains two logical atomic statements:

1. "Diffusion is a form of passive transport."
2. "Osmosis is a form of passive transport."

Some sentences were quite complex and contained more than two logical atomic statements, for example:

> "Simple diffusion involves small molecules passively moving through a membrane from areas of high to low concentration."

This sentence contains four logical atomic statements:

1. "Simple diffusion involves small molecules."
2. "Simple diffusion involves passive movement."
3. "Simple diffusion involves a membrane."
4. "Simple diffusion involves movement from areas of high to low concentration."

We decided to analyze responses as sets of correct and/or incorrect statements (versus whole sentences) for two main reasons. First, we thought it would simplify the veracity coding process (described later). In fact, we defined logical atomic statements as statements that are entirely true or entirely false. Second, we wanted to give participants partial credit for sentences with some true statements and some false statements. Consider the following example:

"Diffusion is a passive process, except with facilitated diffusion where it requires energy."

This sentence contains 2 statements: (1) diffusion is a passive process and (2) facilitated diffusion requires energy. The first statement is true and the second statement is false.

The process of splitting sentences into statements was validated as follows. First, after developing an initial coding guide (Appendix B), researches R1 and R2 coded all open-ended responses from 4 participants (10% of responses). During this process, responses were sentence-segmented and both researchers independently split each sentence into logical atomic statements. After this, R1 manually assigned a Statement ID to each statement identified by R1 and R2. This manual process was validated later through a different exercise. Assigning a Statement ID to each statement enabled us to consider the agreement between the statements identified by R1 and R2. Naturally, two statements might be written differently by R1 and R2 but be semantically equivalent and therefore have the same Statement ID. Consider the following two statements identified by R1 and R2:

- **R1** potential fact: "Osmosis requires a semipermeable membrane"
- **R2** potential fact: "Osmosis needs a semipermeable membrane"

Both statements were assigned the same Statement ID of osm_semiperm because they are semantically equivalent.

To measure agreement, we measured the Jaccard coefficient between the Statement IDs of statements identified by R1 and R2 for each response, using the following formula:

$$J(R_1, R_2) = \frac{|R_1 \cap R_2|}{|R_1 \cup R_2|} \tag{4.2}$$

where R_1 is the set of statements identified in sentences by R1 and R_2 is the set of statements identified in sentences by R2.

After the first round of coding, the Jaccard coefficient (averaged across responses) was 0.56. We deemed this value too low. Therefore, R1 and R2 discussed their disagreements and revised the coding guide. Next, R1 and R2 repeated the same process with a different set of 4 participants (a different 10% of the data). After this second round of coding, the Jaccard coefficient (averaged across responses) was 0.83. We deemed this value to be high enough to establish the reliability of the process of identifying logical atomic statements within a sentence. R1 was therefore responsible for:

(1) splitting the remaining 90% of the data into statements (including the original 10% for which agreement was low) and (2) assigning a Statement ID to each statement.

4.10.2 Assigning Statement IDs to Statements

A post hoc analysis was then conducted to verify the assigning (by R1) of Statement IDs to statements. The process of assigning Statement IDs to statements was validated as follows. First, R1 developed a small coding guide (Appendix C). Then, R1 compiled a list of all Statement IDs across all 80 open-ended responses (i.e., 100% of the data) and a selected representative statement for each Statement ID. Across all open-ended responses there were 374 Statement IDs (i.e., semantically distinct statements). Finally, R2 analyzed all statements from four participants (10% of the data) and assigned each statement a Statement ID. The Cohen's Kappa agreement between R1 and R2 was $\kappa = 0.982$, which is considered "almost perfect" agreement [145].Given this high level of agreement, R1 was solely responsible for assigning Statement IDs to all statements across all participants.

4.10.3 Veracity of Statements

Finally, the veracity of the statements were assessed. Two expert reviewers (professors of Biology at Southern Arkansas University) assessed each of the 374 unique statements as true or false. Both researchers reviewed all statements concurrently and agreed upon whether each statement was true or false.

If a statement was deemed false, the expert reviewers were asked to provide a short justification. Such justifications took different forms. First, false statements could simply be judged as false because the opposite is true. For example, the statement "facilitated diffusion requires energy" was coded as false with the justification "Facilitated diffusion does not require energy." Second, statements might have been deemed false because a word or concept was not correctly remembered. As an example, "Hypotropic indicates low concentration" was marked as false with the justification "Hypotonic [not hypotropic] indicates low concentration." Third, statements with an overstated claim or incorrect premise were deemed false. For example, "Particles naturally want to move from areas of high to low concentration." was marked as false with the following justification, "Particles don't 'want' anything." Fourth, some statements remembered relevant descriptive words but associated them to incorrect concepts. As an example, "Active diffusion involves ATP." was marked as false with the justification, "Active transport involves ATP." The assessment resulted in 312 true statements and 62 false statements.

4.11 Think-Aloud and Search Behaviors

To address **RQ3** and **RQ4**, I captured think-aloud data and search behaviors. Think-aloud data was captured via Zoom recording. Zoom recordings also captured participant screen activities associated with or without verbalization (e.g., copy/pasting text into the Subgoal Manager, highlighting portions of a document with cursor).

Search behaviors were also captured on the SERP and the auxiliary search tools (i.e., Subgoal Manager and Text Editor). In my analysis, each of the SERP items captured are considered independently. However, in order to better understand the significance of each item in terms of participant search behavior they have been grouped and detailed as follows:

- **Query Characteristics**
  - *Queries*: number of all queries issued by a participant.
  - *Distinct Queries*: number of unique queries issued by a participant.
  - *Average Query Length*: average query length in words.
  - *Queries with Question Words*: queries that contained a question word of "who", "what", "where", "why", "when", or "how".
- **Query Abandonment**
  - *Queries without Scrolls*: number of queries issued without a subsequent scroll event.
  - *Queries without Mouseovers*: number of queries issued without a subsequent mouseover event.
  - *Queries without Clicks*: number of queries issued without a subsequent click event.
  - *Queries with Quick Reformulation*: number of queries issued $\leq$ 30 seconds from the previous query.
- **SERP Click Characteristics**
  - *Clicks*: number of SERP clicks.
  - *Web Clicks*: number of SERP clicks within the main Web tab.
  - *Non-Web Clicks*: number of SERP clicks within the Images, Video, and News tabs.

- *Distinct URLs*: number of unique SERP results clicked by a participant.
- *Average Click Rank*: average rank across all search results clicked.

- **Pace of Interaction**
  - *Time Between Events*: average time (in seconds) between subsequent events: queries and clicks.
  - *Search Completion Time*: time difference in minutes between the first query and exiting the search system.
  - *Queries per Minute*: rate of queries calculated from dividing *Queries* by *Search Completion Time*.
  - *Clicks per Minute*: rate of clicks calculated from dividing *Clicks* by *Search Completion Time*.

- **Singularity of Interaction**
  - *Unique Queries*: number of queries not issued by any other participant.
  - *Unique Query Terms*: number of query terms (i.e., words) not issued by any other participant.
  - *Unique URLs*: number of clicked URLs not clicked by any other participant.

The first four items capture each participant's query characteristics and include number of total queries and average length of query. The next four items relate to query abandonment by capturing events like queries without subsequent scrolling and queries issued within 30 seconds from the last query. The next five items capture the characteristics of clicks made on the SERP. These items involve number of clicks overall, where the clicks took place (e.g., Web tab versus Images tabs), and the number of unique SERP results that were clicked by a particular participant. The next set of four items relate to a participant's pace of interaction. Such items capture number of clicks per minute, queries per minute, and the average time a participant spent between events (i.e., query and click events). The last three items capture a participant's singularity or uniqueness of interaction. These items include the number of queries, query terms, and SERP results that were only issued or selected by one particular participant.

The content of the Subgoal Manager and Text Editor tools were also captured. In the Subgoals condition, I captured the content of each subgoal, all notes taken for specific subgoals, all instances in which subgoals were marked as completed, and deletions of subgoals.

4.12 Qualitative Coding of SRL Processes

In **RQ4**, I investigate the effects of the subgoal condition on the SRL processes that participants engaged with during the learning-oriented search task. To address **RQ4**, we conducted a qualitative analysis of participants' search sessions by leveraging their recorded think-aloud comments and screen activities, which included all search, note-taking, and reading activities. SRL processes were coded using the SRL Coding Guide shown across Tables 4.11, 4.13, 4.15, and 4.17. The SRL Coding Guide was adapted from Greene et al. [37, 79]. The coding guide consists of four macro-SRL processes of *Planning*, *Strategy Use*, *Monitoring*, and *Interest*. Three of the macro-SRL processes contain multiple micro-SRL processes: *Planning* contains 6 micro-SRL processes; *Strategy Use* contains 16 micro-SRL processes; and *Monitoring* contains 14 micro-SRL processes. Tables 4.11, 4.13, and 4.15 provide the definition and an example of each micro-SRL process. Seven of these micro-SRL processes are new and were created in order to capture SRL behaviors that were relevant in the current study, and search-based studies in general, but not captured in a previously established SRL Coding Guide. Novel micro-SRL processes are marked with (*) in Tables 4.11, 4.13, and 4.15.

Screen recordings and think-aloud comments were categorized by two researchers (R1 and R2) into distinct SRL processes using the following qualitative analysis process. First, R1 segmented 100% of the data into codable units. Codable units were identified as either a behavior (e.g., typing notes, issuing a search query) or a think-aloud comment (e.g., "Now I'm going to focus on how diffusion is different from osmosis.") indicative of a particular micro-SRL process. The data was then coded in three rounds. During each round of coding, R1 and R2 assigned each codable unit to a micro-SRL process using the definitions and examples provided in the SRL Coding Guide. Because each micro-SRL process is associated with a macro-SRL process, each codable unit was effectively associated with both a single micro-SRL process and a single macro-SRL process.

In the first round of coding, R1 and R2 independently coded search sessions from the same 4 participants (10% of participants). In this initial round, after coding each participant R1 and R2 met and resolved disagreements, then modified the SRL Coding Guide to address issues that came up in disagreement resolution discussions. In the second round of coding, R1 and R2 again coded

Table 4.11 Definitions of Planning micro-SRL processes with examples. Micro-SRL processes marked with (*) are novel.

Planning	**Macro-SRL process that involves assessing the task and constraints along with developing, modifying, and revisiting subgoals, and/or organizing subgoals**
Micro-SRL	*Description*
Modifies Subgoals*	Modifies (verbally or through typing) or deletes existing goal (does not include cases of editing while initially establishing subgoal).
	[Edits subgoal “Find 3 examples of osmosis in everyday life” to “Find 2 examples of osmosis in everyday life”] OR [Deletes subgoal]
Planning	Sets multiple subgoals or one multi-component subgoal.
	“First I’ll look up the definition of diffusion, then I’ll look up a few examples.”
Recycle Goal in Working Memory	Re-states current subgoal.
	“Ok, what am I doing? I need to understand how diffusion is different from osmosis”
Revisits Previous Subgoal*	Engages in different, already established subgoal.
	[clicks on a different subgoal than they were currently pursuing in the SM and takes action toward subgoal] OR "I think I’m going to go back to diffusion."
Revisits task*	(Re-)reads task description.
	[Opens Task Description page and reads description]
Subgoals	Sets new subgoal toward overall goal and then takes action in response.
	[Writes “Diffusion” in TE and then queries “Diffusion definition”.]

Table 4.13 Definitions of Strategy Use micro-SRL processes with examples. Micro-SRL processes marked with (*) are novel.

Strategy Use	**Macro-SRL process that involves using a study strategy or tactic toward a goal that involves some product (e.g., notes) or source (e.g., webpage, prior knowledge).**
Micro-SRL	*Description*
Comparing & Contrasting	Compares/contrasts two externalized representations or ideas.
	"It says high to low" [opens different subgoal to compare to notes]
Copying Notes	Copy/pastes information.
	[Copy and pastes into TE/SM]
Corroborating sources	Compares information from different sources (a source can be notes taken from a prior source) to evaluate accuracy.
	"That's weird because we read that it is low to high concentration."
Draw	Makes drawing.
	[Types picture of osmosis flow through semipermeable membrane out of symbols into TE]
Emphasizing notes*	Underlines, bolds, capitalizes or otherwise emphasizes text after noting its importance, i.e., reformat to increase visual salience.
	[Bolds the word "semipermeable" in definition of osmosis]
Forming New Conclusion*	Draws a conclusion based on information from one or more sources; includes *Knowledge Elaboration*, *Hypothesizing*, *Inferences*
	"So if the solvent is water, then the water would move."
Help-seeking behavior	Desires help with something and asks the moderator.
	"Are we allowed to write the stuff in our text boxes here?"
Manipulate representation	Controls visual representation (e.g., video pause, graphic zoom)
	[Pauses YouTube video]
Memorization	Tries to memorize verbatim (e.g., repeats information aloud more than once, closes webpage then tries to restate what was read)
	"So, osmosis is the movement of solvent particles...so osmosis is the movement of solvent particles across a semipermeable membrane"
Prior knowledge activation	Takes inventory of prior knowledge in order to: (1) develop subgoals; (2) pursue new subgoal; or (3) take down notes of what is already known with respect to a particular subgoal
	"I remember the practice test asking about...dye and different versions of solutes."
Reading notes	Reads aloud something already written or copy/pasted
	[Reads notes aloud]
Re-reading	Re-reads aloud something not written by him/her
	[Re-reads paragraph in web page]
Search	Issues a new query (not reformulation)
	[Issues query for "diffusion definition biology"]
Select new informational source	Navigates to new webpage (not subsequent visits)
	[Clicks on SERP result]
Self-knowledge activation	Statement to pursue or avoid strategy based on personal preference/aptitude (looking forward)
	"We're going to look at pictures because I'm a big pictures person"
Taking Notes	Writes notes either word-for-word or own words.
	[Types in TE/SM]

Table 4.15 Definitions of Monitoring micro-SRL processes with examples. Micro-SRL processes marked with (*) are novel.

Monitoring	**Macro-SRL process that involves comparing products (e.g., notes, drawing) with standards (i.e., those set in subgoals) to see if they align or evaluating sources (e.g., webpage, prior knowledge) to see if they help in achieving standards.**
Micro-SRL	*Description*
Content Evaluation	assesses relevance based on goal or subgoal (e.g., states that information is or is not useful).
	"I'm not going to include it because it is too vague"
Expectation of adequacy of content	estimates usefulness before reading carefully (e.g., review SERP snippets and then query reformulation).
	"It goes into too much detail"
Feeling of Knowing (FOK)	feels that they know but is unable to retrieve on demand.
	"What's it called? Something chain?"
Feeling of Recognition (FOR)	States that they do or do not know something (e.g., "I know this", "I have learned about this before", "I have no idea what this is").
	"I've definitely studied this before"
Judgment of Learning (JOL)	States that they have or have not learned something well enough for future use (e.g., "Ok, I think I understand this well enough to answer questions about it") OR tests or checks learning with external quiz questions.
	"If I were asked a question like the other one I would be able to do it"
Judgment of Understanding (JOU)	States that they understand, think they understand, or do not understand (e.g., "this makes sense", "I don't understand") OR checks/confirms understanding (e.g., "that's what I thought")
	"I still don't get what decides if it swells or not"
Monitor progress toward subgoals	assesses progress toward goal(s).
	"I feel like that meets everything."
Monitor subgoal quality*	Assess the quality of a goal (e.g., content, relevance, usefulness)
	"I think it will take me a bit of time to find these examples, so I think these are good goals."
Monitor use of strategies	Evaluates the usefulness of a strategy (only looking back)
	[After taking an online quiz]; "Can I do it again because that was helpful"
Questioning Task Expectation*	Questions the expectations of the task
	"Oh, what type of biology exam is this?"
Representation difficulty	States that representation is easy/difficult
	"A heavier text like this I don't enjoy looking at because I like to see some pictures, some explanation."
Self-Questioning	States unknown without forming a plan (must involve a question mark) (e.g., "So are solvents involved in hypertonic solutions?")
	"How is the liquid higher on that side?"
Task difficulty	Expresses that the (sub)task is easy/difficult
	"This stuff is hard to even, like, read."
Time monitoring	Checks remaining time OR references time with respect to goals
	"Looking at the time I'm changing my mind."

Table 4.17 Coding scheme for analysis of Interest micro-SRL processes.

Interest	**Macro-SRL process that involves expressing dis/interest in task-related information**
Micro-SRL	*Description*
Interest Statement	Expresses that something task-related is interesting/boring not including instances where participants use interesting as a synonym of peculiar (e.g., "I find it interesting they added this information about a different topic in this article"), but rather interest as in fascinating or captivating (e.g., "Wow! This is so interesting, osmosis helps us to breathe!!").
	"Oh so it can happen in gases too. That's pretty cool!"

four participants. After coding all four participants in isolation, the Cohen's Kappa agreement was $\kappa = 0.872$ between micro-SRL processes and $\kappa = 0.944$ between macro-SRL processes, both of which are considered "almost perfect" agreement [145]. R1 then re-coded the initial 4 participants from the first round and an additional 24 participants for a total of 32 participants (80% of participants). Then, a final round of coding was conducted in order to ensure there had not been any coding drift since the second round of coding. In the final round of coding, R1 and R2 each coded two participants (5% of participants). The Cohen's Kappa agreement was $\kappa = 0.919$ between micro-SRL processes and $\kappa = 0.944$ between macro-SRL processes, both of which are considered "almost perfect" agreement [145]. Because agreement was high, R1 coded the final six participants.

4.13 Statistical Analysis

In order to detect potential differences between subgoal conditions, I selected a non-parametric test because it is unreasonable to assume normality in the population. Normality requires unbounded data. The variables of interest in this study are bounded by zero (e.g., SRL frequency from think-aloud and screen recordings) or bounded on both ends (e.g., likert scale questionnaire items). Further, the boundary is frequently met (e.g., reaches zero). Therefore, it is unreasonable to assume that the population follows a normal distribution. I use the non-parametric Wilcoxon Rank Sum Test (also known as the Mann-Whitney U Test) for this analysis. I selected this test as it is suitable for detecting differences between independent groups (i.e., suitable for this study's between subjects design in which participants were only exposed to a single condition).

Additionally, the Shapiro Wilk test for normality was conducted for all 68 variables of analysis and, of those, only 9 variables were likely to follow a normal distribution, while the remaining 59 were likely to follow a non-normal distribution. Those 9 variables were: 3 pre-task questionnaire variables (Perceptions of Prior Knowledge; Expected Difficulty; and A Priori Determinability); 2 post-task questionnaire variables (Perceptions of SRL Adapting and Focused Attention); 1 learning assessment variable (Percent of True Facts on the open-ended post-task learning assessment); and 3 SRL variables (Strategy Use Macro-SRL; Summarization Micro-SRL; and Diversity of Strategy Use Micro-SRLs). None of these variables were significantly different between groups under a T-Test nor Wilcoxon Rank Sum Test.

A two-tailed Wilcoxon Rank Sum Test involves the following hypotheses:

- Null hypothesis (H_0): The two populations are equal.
- Alternative hypothesis (H_1): The two populations are not equal.

A one-tailed Wilcoxon Rank Sum Test involves the following hypotheses:

- Null hypothesis (H_0): The two populations are equal.
- Alternative hypothesis (H_1): One population is less than the second population.

The variables associated with learning outcomes and SRL have specific hypotheses (described in Chapter 3). I have conducted one-tailed tests for these variables. This is because the research hypotheses are testing for differences in a specific direction (i.e., that the NoSubgoals condition will be less than the Subgoals condition). The variables associated with interest, perceived knowledge increase, expected difficulty, satisfaction, engagement, and all search behavior variables are exploratory and do not have specific directional research hypothoses. For this reason, I have conducted two-tailed tests for these variables.

Given two groups X and Y both of size n, the Wilcoxon Rank Sum test is calculated in several steps. First, both groups X and Y are combined, then sorted in ascending order. Next, the data is assigned a rank based on that order. Then, the data is separated back into X and Y groups maintaining their assigned ranks. Next, the ranks are summed for each group X and Y giving a W_X and W_Y (i.e., sum of the ranks) for each group. Each W is used to calculate a test statistic U for each group:

$$U = W - \frac{n(n+1)}{2} \tag{4.3}$$

where n is the sample size of the group. U is calculated for both groups X and Y, then the minimum is reported as the final U statistic.

CHAPTER 5

Results

In the following sections, I report on results for research questions **RQ1–RQ4**. All research questions involved comparisons between subgoal conditions. To test for statistically significant differences between groups (i.e., NoSubgoals and Subgoals), I used the Wilcoxon Rank Sum Test also known as the Mann-Whitney U Test (detailed in Section 4.13).

First, I present results from the pre-task questionnaire and pre-task multiple-choice ODCA learning assessment. These were collected to ensure there were no significant differences between groups before the search session.

5.1 Differences in Pre-Task Perceptions & Prior Knowledge

Here, I present results from the pre-task questionnaire and pre-task multiple-choice ODCA learning assessment. These variables were captured to ensure there were no significant differences between groups in terms of task interest, expected difficulty, and a priori determinability. Self-reported prior knowledge and scores on the pre-task ODCA assessment were compared to ensure there were no significant differences between groups in terms of prior knowledge.

Figures 5.1-5.5 show differences in pre-task perceptions between groups. There were no statistically significant differences detected between groups in terms of: (1) interest ($U = 180$, $p = 0.59$); (2) prior knowledge ($U = 136$, $p = 0.09$); (3) expected difficulty ($U = 241$, $p = 0.27$); (4) a priori determinability ($U = 194$, $p = 0.88$); nor (5) percentage score on pre-task ODCA ($U = 217$, $p = 0.14$). Table 5.1 shows the median, maximum, minimum, and standard deviation of pre-task perceptions and pre-task ODCA percentage scores for each group.

5.2 Differences in Learning

In this section, I present results addressing **RQ1** associated with learning outcomes. In my study, learning was measured with a multiple-choice assessment called the ODCA at three points in time (i.e., pre-task, post-task, and retention) and with an open-ended assessment at two points in time (i.e., post-task and retention). Results from the pre-task assessment captured prior knowledge and

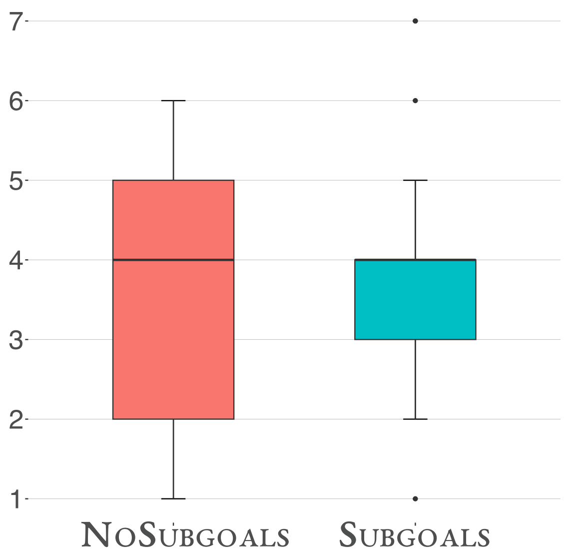

Figure 5.1 Pre-task interest

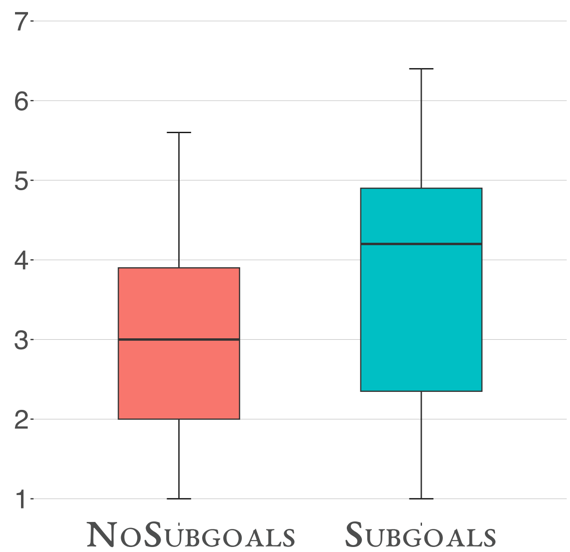

Figure 5.2 Pre-task prior knowledge

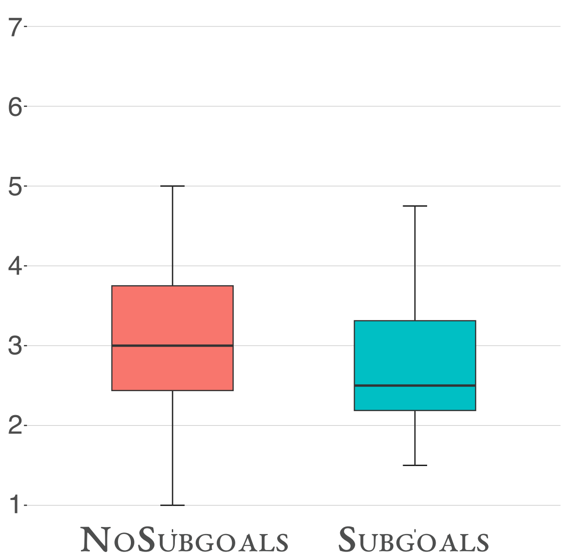

Figure 5.3 Pre-task expected difficulty

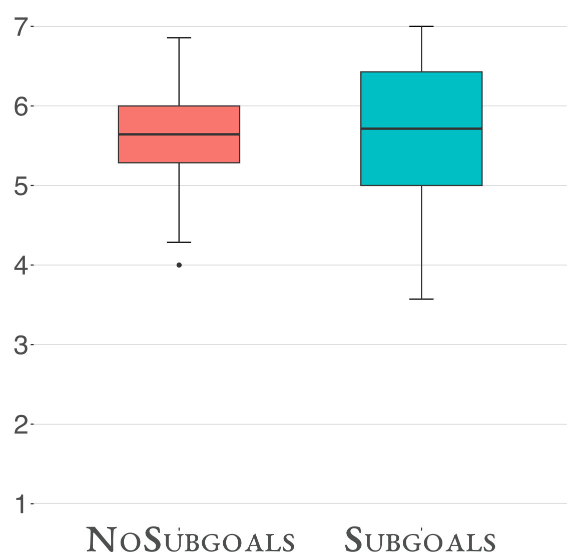

Figure 5.4 Pre-task a priori determinability

were reported above. All post-task and retention assessments are presented in this section.

First, I present the results from the multiple-choice ODCA assessments. I provide differences between groups in terms of: (1) the normalized gain on the post-task assessment; and (2) the normalized gain on the retention assessment.

Second, I present the results from the open-ended assessments. I provide differences between groups in terms of: (1) the percent of true statements on the post-task assessment; (2) the percent of true statements on the retention assessment; and (3) the percent of true statements retained from

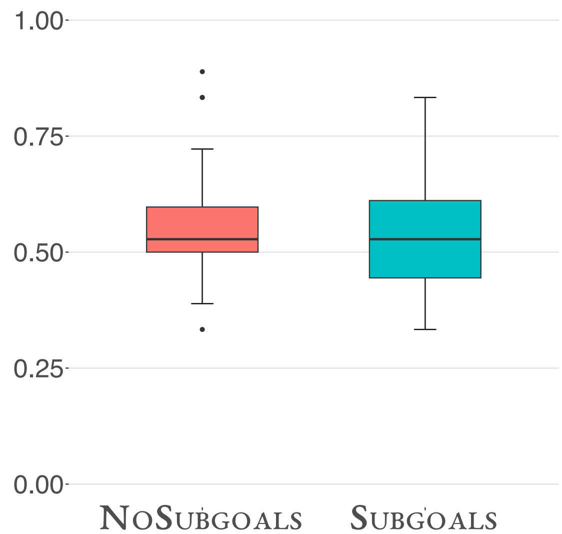

Figure 5.5 Percentage Score on Pre-Task ODCA

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Interest	4.000	6.000	1.000	1.468	4.000	7.000	1.000	1.531
Reported Prior Knowledge	3.000	5.600	1.000	1.329	4.200	6.400	1.000	1.633
Exepected Difficulty	3.000	5.000	1.000	1.105	2.500	4.750	1.500	0.831
A Priori Determinability	5.643	6.857	4.000	0.690	5.714	7.000	3.571	0.955
Pre-Task ODCA Percentage	0.528	0.889	0.333	0.155	0.528	0.833	0.333	0.141

Table 5.1 Pre-Task Perceptions & Prior Knowledge Descriptive Statistics

the post-task assessment to the retention assessment.

5.2.1 Multiple-Choice ODCA Assessment

Figure 5.6 show differences in normalized gain on the post-task ODCA between conditions. On average, participants scored higher in the Subgoals versus NoSubgoals condition. However, there was no statistically significant difference detected between conditions ($U = 150$, $p = 0.088$).

Figure 5.7 show differences in normalized gain on the retention ODCA between conditions. On average, participants scored higher in the Subgoals versus NoSubgoals condition. There was a statistically significant difference detected between conditions ($U = 127$, $p < 0.05$).

Table 5.2 shows the median, maximum, minimum, and standard deviation of normalized gain on the post-task ODCA and retention ODCA for each group.

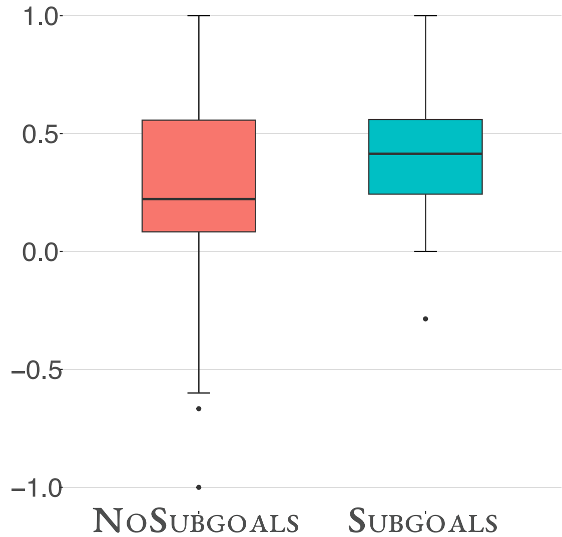

Figure 5.6 Normalized Gain on Post-Task ODCA

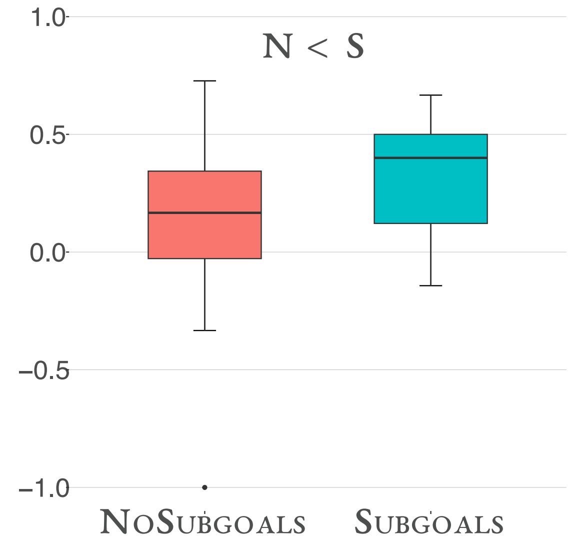

Figure 5.7 Normalized Gain on Retention ODCA

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Post. Norm. Gain	0.222	1.000	-1.000	0.534	0.414	1.000	-0.286	0.309
Ret. Norm. Gain*	**0.167**	**0.727**	**-1.400**	**0.521**	**0.400**	**0.667**	**-0.143**	**0.245**

Table 5.2 Multiple-Choice ODCA Assessment Descriptive Statistics (* indicates statistically significant differences between groups)

5.2.2 Open-Ended Assessment

Figure 5.8 shows differences in the percent of true statements on the post-task open-ended assessment between conditions. On average, participants scored higher in the Subgoals vs. NoSubgoals condition. However, there was no statistically significant difference detected between conditions ($U = 146$, $p = 0.147$).

Figures 5.9 and 5.10 show differences in the percent of true statements on the open-ended retention assessment and the percent of true statements retained across open-ended assessments between conditions. On average, participants included more true statements *and* retained more true statements in the Subgoals vs. NoSubgoals condition. There were statistically significant differences detected between conditions in terms of the percentage of true statements included in the open-ended retention test ($U = 118$, $p < 0.05$) and the percentage of true statements retained ($U = 77$, $p < 0.001$).

Table 5.3 shows the median, maximum, minimum, and standard deviation of the percent of true statements on post-task, the percent of true statements on retention, and the percent of true statements retained across open-ended assessments for each group.

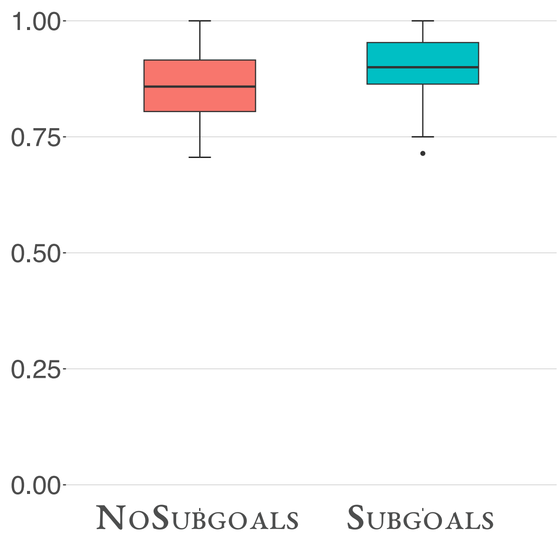

Figure 5.8 Percent of True Statements on Post-Task Open-Ended Assessment

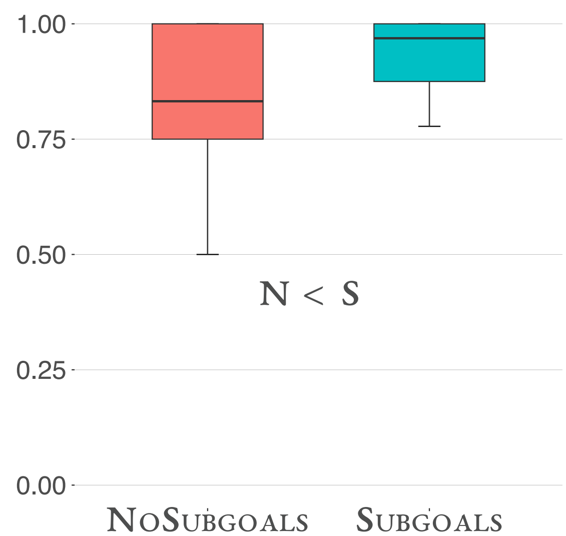

Figure 5.9 Percent of True Statements on Retention Open-Ended Assessment

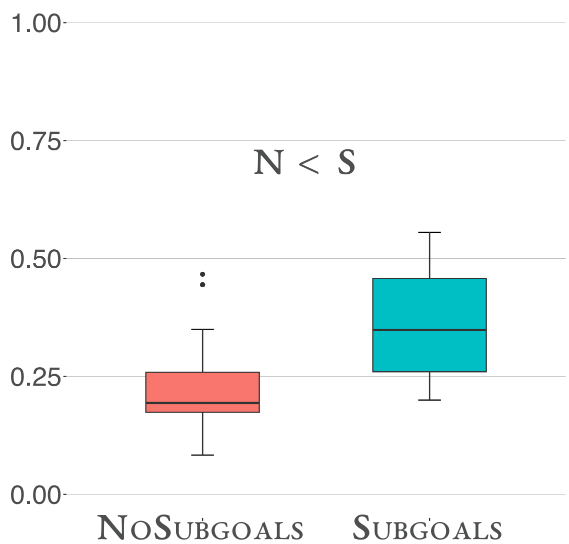

Figure 5.10 Percent of True Statements Retained across Open-Ended Assessments

5.3 Differences in Searcher Perceptions

In this section, I present results addressing **RQ2** associated with searcher perceptions. In my study, searcher perceptions were measured with a pre-task questionnaire and a post-task questionnaire. However, only post-task perceptions were considered in **RQ2** as pre-task perceptions (discussed

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Pct. TS Post.	0.858	1.000	0.706	0.085	0.900	1.000	0.714	0.079
Pct. TS Retention*	**0.832**	**1.000**	**0.500**	**0.136**	**0.969**	**1.000**	**-0.778**	**0.075**
Pct. TS Retained*	**0.194**	**0.467**	**0.083**	**0.112**	**0.348**	**0.556**	**0.200**	**0.115**

Table 5.3 Open-Ended Assessment Descriptive Statistics (* indicates statistically significant differences between groups)

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Interest Inc.	5.000	6.000	1.000	1.468	5.000	7.000	1.000	1.451
Knowledge Inc.	6.500	7.000	4.600	0.674	6.300	7.000	-4.400	0.865
Satisfaction	6.125	7.000	4.000	0.829	5.875	7.000	3.500	1.103

Table 5.4 Post-Task Perceptions Descriptive Statistics

above) were used as a check to ensure there were no significant differences between groups before beginning the search session.

Here, I present the results from the post-task questionnaire. I provide differences between groups of perceptions of: (1) interest increase; (2) knowledge increase; (3) satisfaction; (4) task difficulty; (5) search difficulty; (6) difficulty integrating information; (7) difficulty deciding when one had enough information; (8) *Planning* SRL support; (9) *Strategy Use* SRL support; (10) *Monitoring* SRL support; (11) *Evaluating Progress* SRL support; (12) *Adapting* SRL support; (13) perceived usability; (14) reward; and (15) focused attention.

5.3.1 Perceptions of Interest Increase, Knowledge Increase, and Satisfaction

Figures 5.11-5.13 show differences in post-task perceptions between groups. There were no statistically significant differences detected between groups in terms of: (1) interest increase ($U = 160$, $p = 0.262$); (2) knowledge increase ($U = 232$, $p = 0.390$); nor (3) satisfaction ($U = 224$, $p = 0.531$). Table 5.4 shows the median, maximum, minimum, and standard deviation of post-task perceptions for each group.

5.3.2 Perceptions of Difficulty

Figures 5.14-5.17 show differences in post-task perceptions of difficulty between groups. There were no statistically significant differences detected between groups in terms of: (1) task difficulty ($U = 181$, $p = 0.58$); (2) search difficulty ($U = 201$, $p = 0.988$); (3) difficulty integrating information ($U = 225$, $p = 0.469$); nor (4) difficulty deciding when one had enough information ($U = 250$,

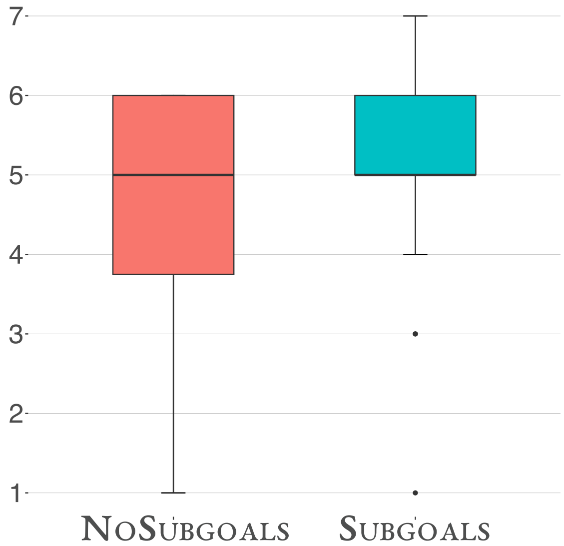

Figure 5.11 Post-task interest increase

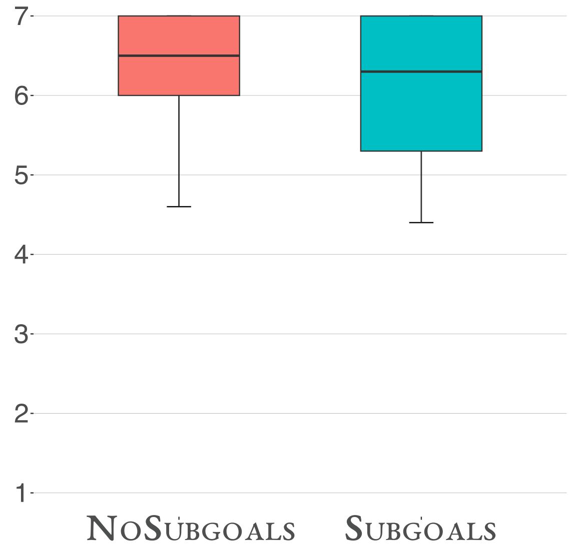

Figure 5.12 Post-task knowledge increase

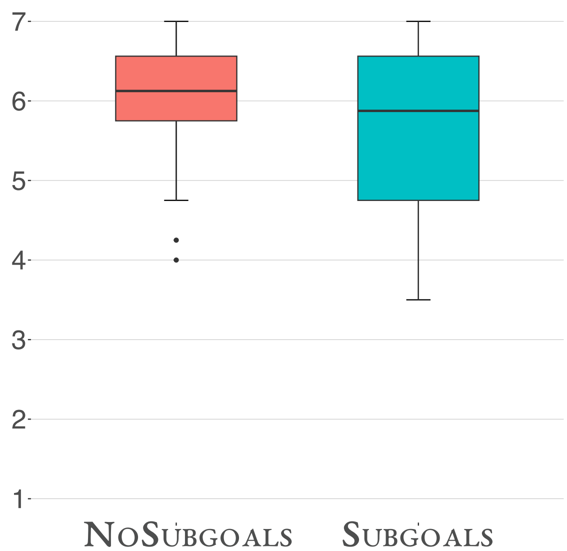

Figure 5.13 Post-task satisfaction

$p = 0.174$). Table 5.5 shows the median, maximum, minimum, and standard deviation of post-task perceptions of difficulty for each group.

5.3.3 Perceptions of SRL Support

Figures 5.18-5.22 show differences in post-task perceptions of SRL between groups. There were no statistically significant differences detected between groups in terms of: (1) *Strategy Use* ($U = 223$, $p = 0.747$); (2) *Monitoring*($U = 168$, $p = 0.190$); nor (3) *Adapting* ($U = 176$, $p = 0.262$).

There were statistically significant differences detected between groups in terms of: (1) *Planning*

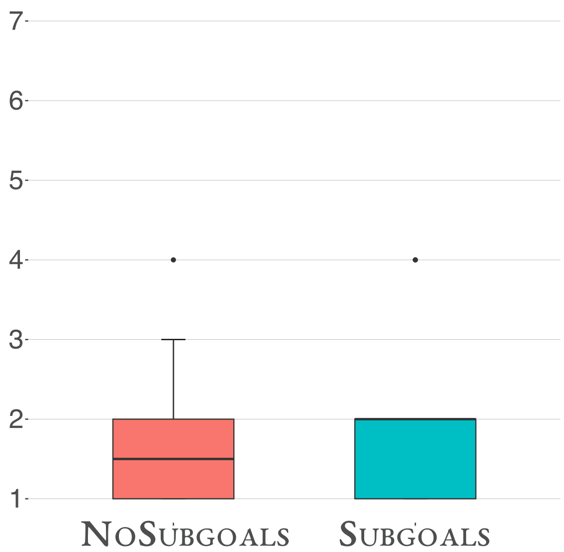

Figure 5.14 Post-task task difficulty

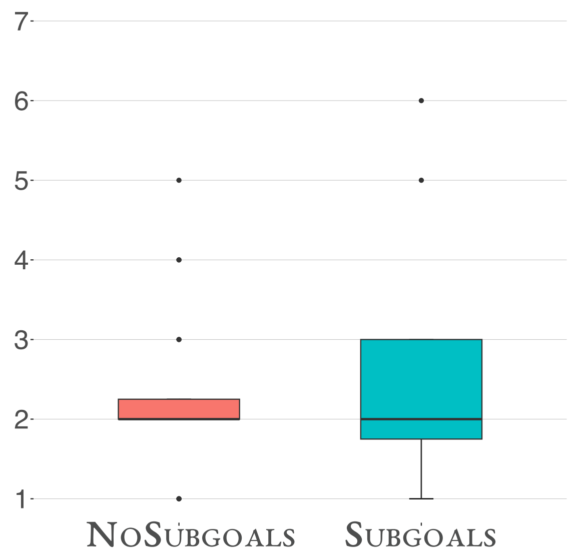

Figure 5.15 Post-task search difficulty

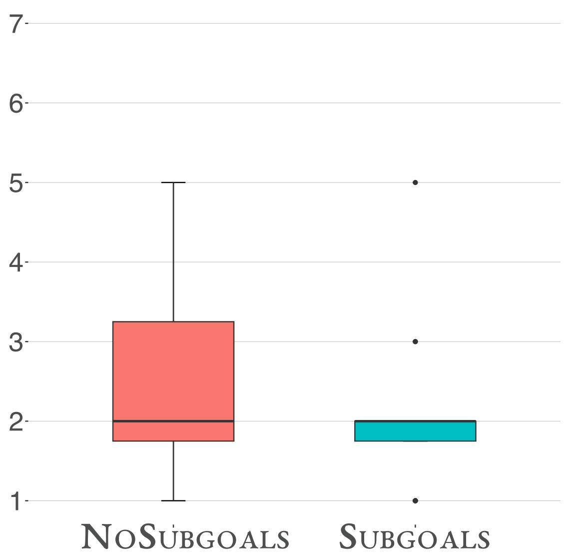

Figure 5.16 Post-task difficulty integrating information

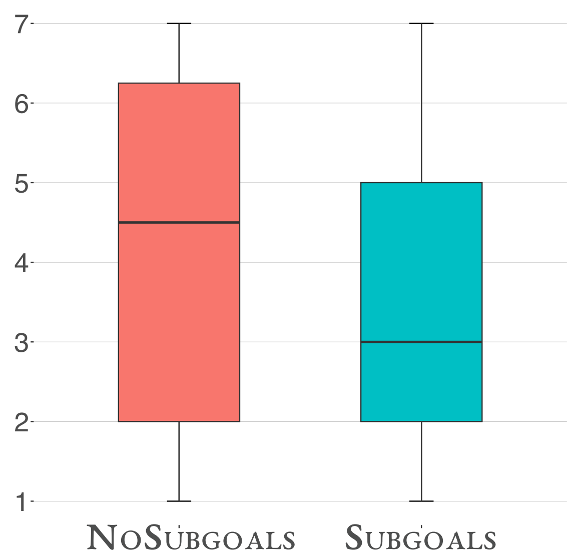

Figure 5.17 Post-task difficulty deciding when one has enough information

($U = 121$, $p < 0.05$) and (2) *Evaluating Progress* ($U = 131$, $p < 0.05$). On average, participants reported higher rates of support for *Planning* and *Evaluating Progress* in the Subgoals versus NoSubgoals condition. Table 5.6 shows the median, maximum, minimum, and standard deviation of post-task perceptions of SRL support for each group.

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Task Diff.	1.500	4.000	1.000	0.813	2.000	4.000	1.000	0.894
Search Diff.	2.000	5.000	1.000	1.031	2.000	6.000	1.000	1.309
Int. Diff.	2.000	5.000	1.000	1.226	2.000	5.000	1.000	0.918
Decide Diff.	4.500	7.000	1.000	2.138	3.000	7.000	1.000	1.960

Table 5.5 Post-Task Perceptions of Difficulty Descriptive Statistics

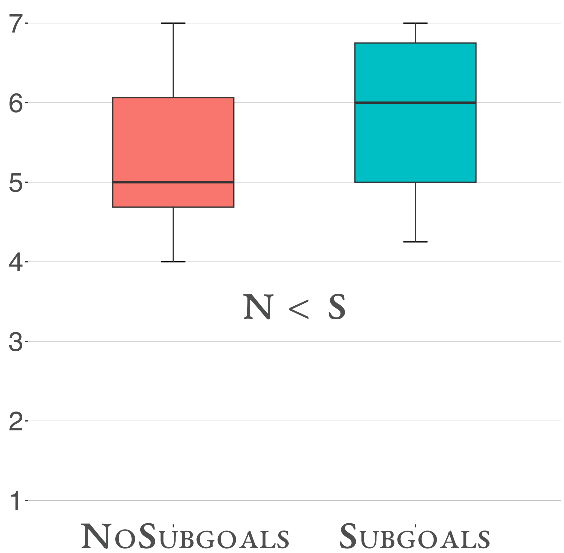

Figure 5.18 Post-task perceptions of SRL planning

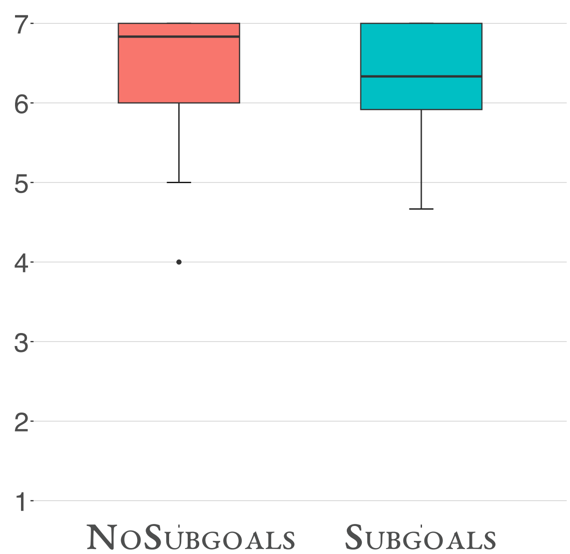

Figure 5.19 Post-task perceptions of SRL strategy use

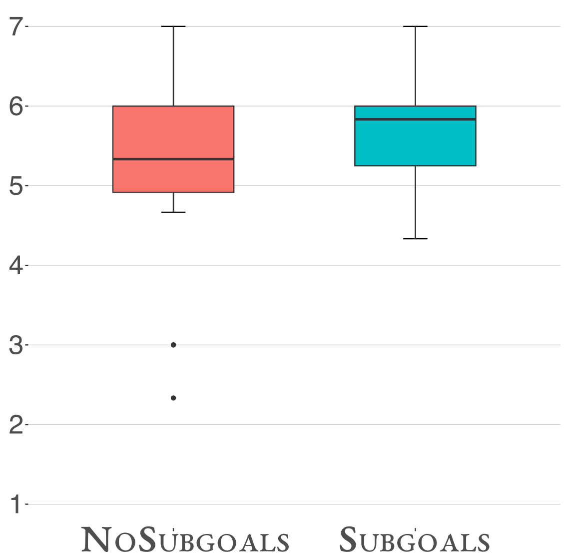

Figure 5.20 Post-task perceptions of SRL monitoring

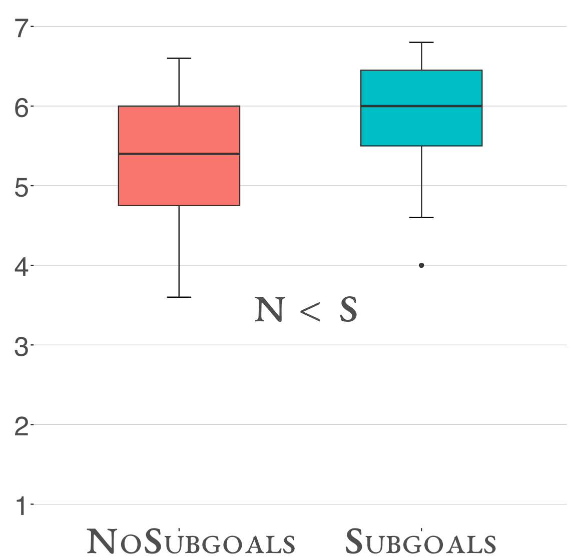

Figure 5.21 Post-task perceptions of SRL evaluating progress

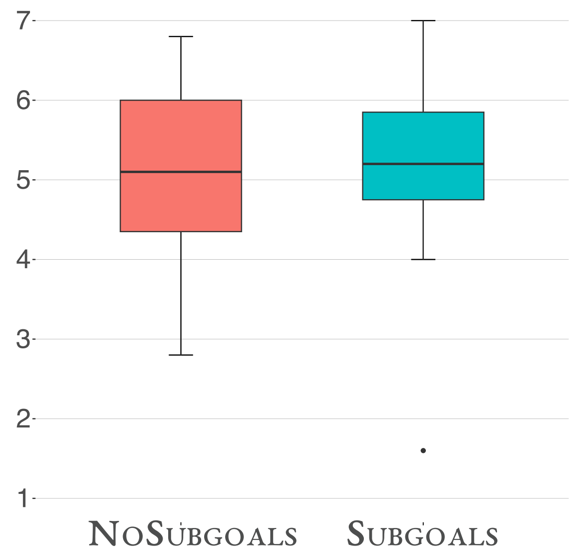

Figure 5.22 Post-task perceptions of SRL adapting

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Planning*	**5.000**	**7.000**	**4.000**	**0.970**	**6.000**	**7.000**	**4.250**	**0.849**
Strategy Use	6.833	7.000	4.000	0.882	6.333	7.000	4.667	0.700
Monitoring	5.333	7.000	2.333	1.265	5.833	7.000	4.333	0.644
Eval. Progress*	**5.400**	**6.600**	**3.600**	**0.884**	**6.000**	**6.800**	**4.000**	**0.787**
Adapting	5.100	6.800	2.800	1.153	5.200	7.000	1.600	1.229

Table 5.6 Post-Task Perceptions of SRL Support Descriptive Statistics (* indicates statistically significant differences between groups)

5.3.4 Perceptions of Engagement

Figures 5.23-5.25 show differences in post-task perceptions of engagement between groups. There were no statistically significant differences detected between groups in terms of: (1) perceived usability ($U = 228$, $p = 0.451$); (2) reward ($U = 220$, $p = 0.595$); nor (3) focused attention ($U = 228$, $p = 0.455$). Table 5.7 shows the median, maximum, minimum, and standard deviation of post-task perceptions of engagement for each group.

5.4 Differences in Search Behavior

In this section, I present results addressing **RQ3** associated with search behavior. In my study, I have developed search behavior variables that are each associated a distinct search behavior dimension. In total, there are five search behavior dimensions: (1) query characteristics; (2) query abandonment;

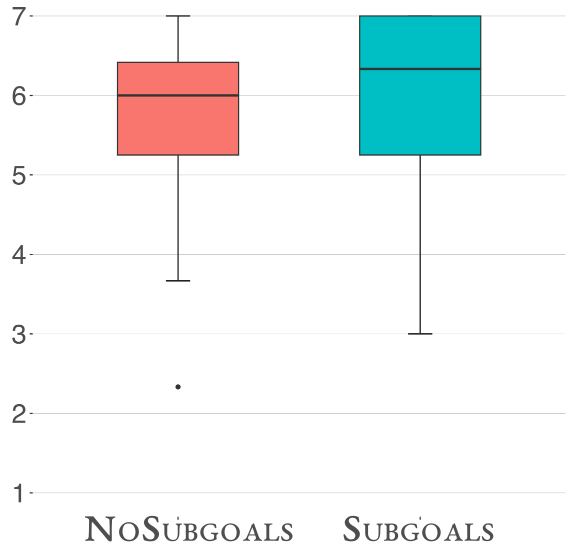

Figure 5.23 Post-task engagement perceived usability

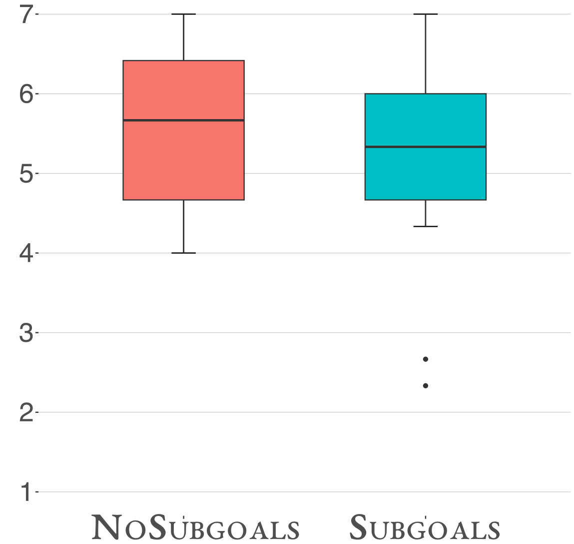

Figure 5.24 Post-task engagement reward

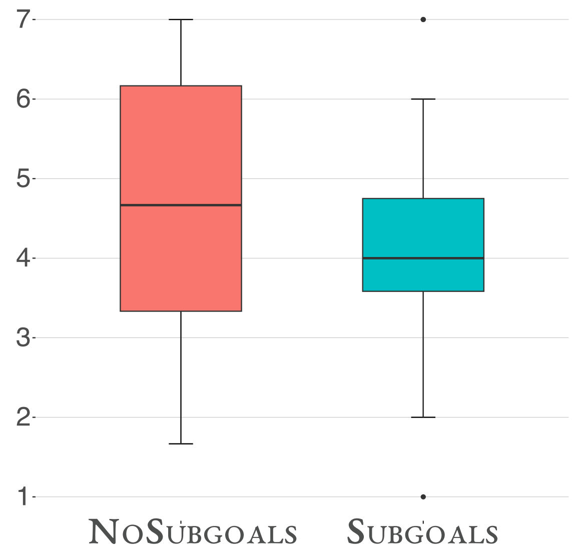

Figure 5.25 Post-task engagement focused attention

(3) SERP click characteristics; (4) pace of interaction; and (5) singularity of interaction.

First, I present the results associated with query characteristics. I provide differences between groups (i.e., NoSubgoals versus Subgoals) of: (1) number of queries; (2) number of distinct queries; (3) average query length; and (4) number of queries with question words.

Second, I present the results associated with query abandonment. I provide differences between groups (i.e., NoSubgoals versus Subgoals) of: (1) number of queries without scrolls; (2) number

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Perceived Usab.	6.000	7.000	2.333	1.207	6.333	7.000	3.000	1.295
Reward	5.667	7.000	4.000	0.988	5.333	7.000	2.333	1.146
Focused Atten.	4.667	7.000	1.667	1.745	4.000	7.000	1.000	1.477

Table 5.7 Post-Task Perceptions of Engagement Descriptive Statistics

of queries without mouseovers; (3) number of queries without clicks; and (4) number of queries with quick reformulation ($<$ 30 seconds).

Third, I present the results associated with SERP click characteristics. I provide differences between groups (i.e., NoSubgoals versus Subgoals) of: (1) number of clicks; (2) number of clicks in the Web tab; (3) number of clicks in the Non-Web tabs; and (4) number of distinct SERP results clicked; and (5) the average rank of SERP results clicked.

Fourth, I present the results associated with pace of interaction. I provide differences between groups (i.e., NoSubgoals versus Subgoals) of: (1) time between events; (2) completion time of search session; (3) number of queries per minute; and (4) number of clicks per minute.

Fifth, I present the results associated with singularity of interaction. I provide differences between groups (i.e., NoSubgoals versus Subgoals) of: (1) number of unique queries; (2) number of unique query terms; and (3) number of unique SERP results clicked.

5.4.1 Query Characteristics

Figures 5.26-5.29 show differences in query characteristics between groups. There were no statistically significant differences detected between groups in terms of: (1) number of queries ($U = 153$, $p = 0.21$); (2) number of distinct queries ($U = 150$, $p = 0.16$); (3) average query length ($U = 136$, $p = 0.09$); nor (4) number of queries with question words ($U = 195$, $p = 0.89$). Table 5.8 shows the median, maximum, minimum, and standard deviation of query characteristics for each group.

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Queries	7.000	21.000	2.000	5.576	9.000	24.000	3.000	5.254
Dist. Queries	7.000	20.000	2.000	5.463	8.500	21.000	3.000	4.628
Avg. Qry. Length	2.955	4.857	1.000	0.976	3.463	5.000	1.800	0.863
Queries w/ Q. Words	0.000	10.000	0.000	2.501	0.000	6.000	0.000	1.954

Table 5.8 Query Characteristics Descriptive Statistics

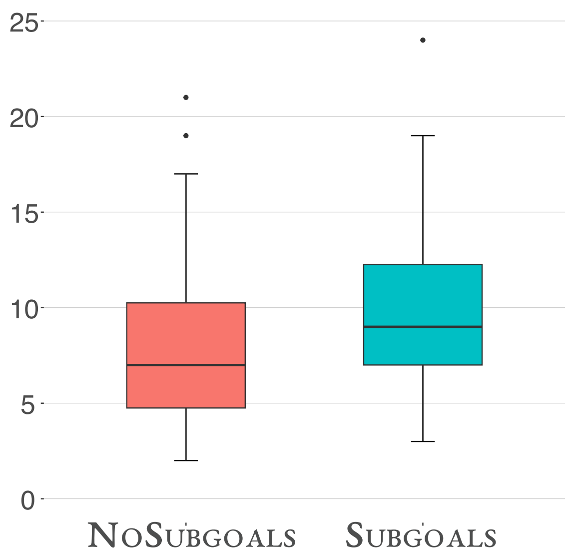

Figure 5.26 Queries (total issued by a participant)

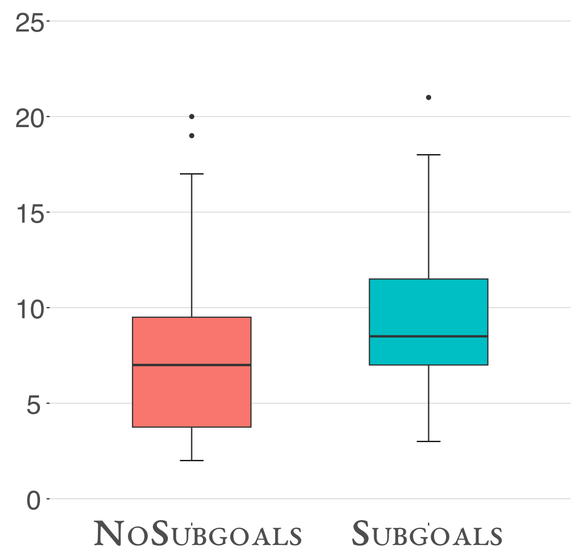

Figure 5.27 Distinct queries (number of unique queries issued by a participant)

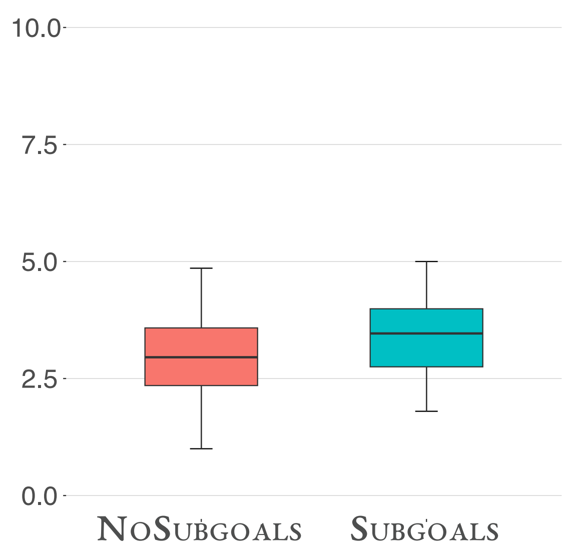

Figure 5.28 Average query length (in words)

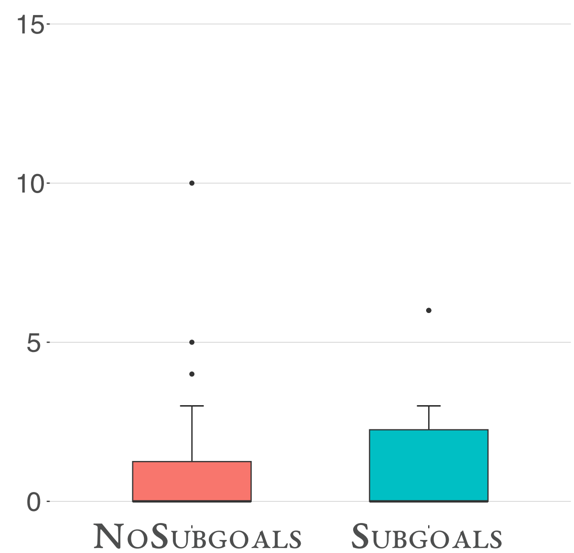

Figure 5.29 Queries with question words (containing who, what, where, why, when, or how)

5.4.2 Query Abandonment

Figures 5.26-5.29 show differences in query abandonment between groups. There were no statistically significant differences detected between groups in terms of: (1) number of queries without scrolls ($U = 185$, $p = 0.68$); (2) number of queries without mouseovers ($U = 181$, $p = 0.56$); (3) number of queries without clicks ($U = 174$, $p = 0.48$); nor (4) number of queries with quick reformulation ($U = 145$, $p = 0.12$). Table 5.9 shows the median, maximum, minimum, and standard deviation of query abandonment for each group.

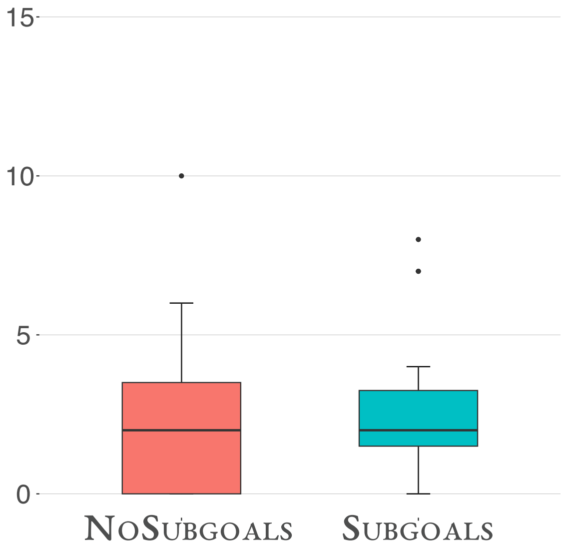

Figure 5.30 Queries without scrolls (number of queries without subsequent scroll event)

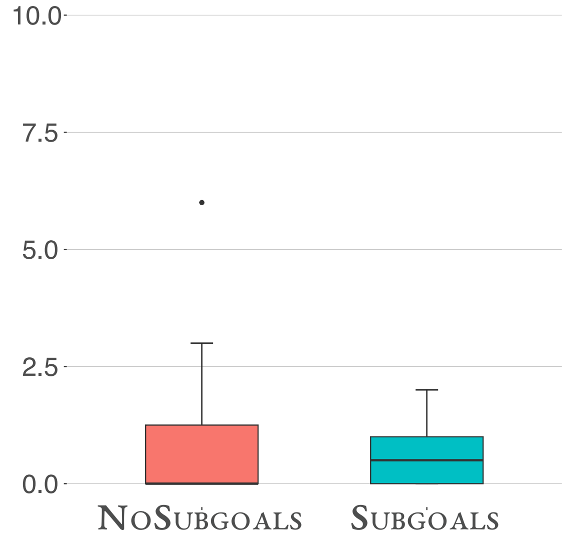

Figure 5.31 Queries without mouseovers (number of queries issued without subsequent mouseover event)

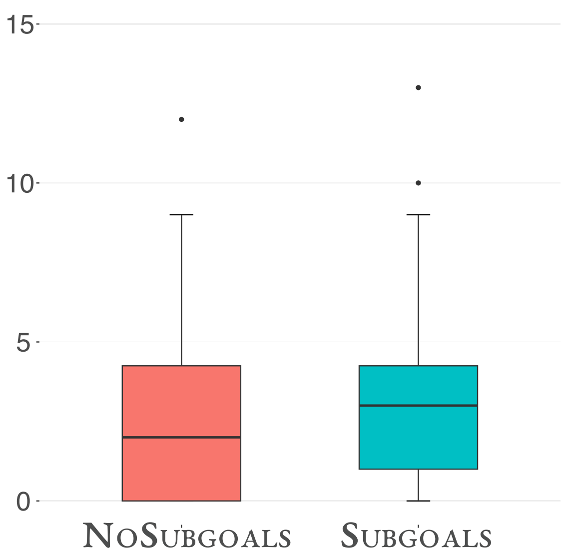

Figure 5.32 Queries without clicks (number of queries issued without subsequent click event)

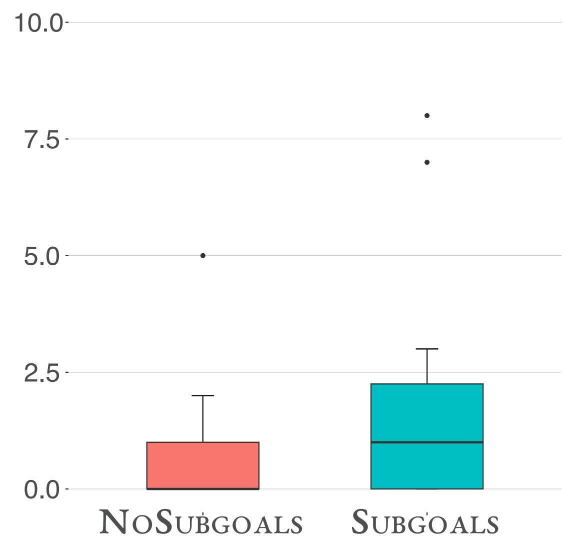

Figure 5.33 Queries with quick reformulation (number of queries issued $\leq$ 30 seconds from the previous query)

5.4.3 SERP Click Characteristics

Figures 5.34-5.38 show differences in SERP click characteristics between groups. There were no statistically significant differences detected between groups in terms of: (1) number of clicks ($U = 146$, $p = 0.15$); (2) number of clicks in the Web tab ($U = 142$, $p = 0.12$); (3) number of clicks

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Queries w/o Scrolls	2.000	10.000	0.000	2.743	2.000	8.000	0.000	2.393
Queries w/o MOs	0.000	6.000	0.000	1.919	0.500	2.000	0.000	0.745
Queries w/o Clicks	2.000	12.000	0.000	3.372	3.000	13.000	0.000	3.543
Qrys. w/ Qck. Reform.	0.000	5.000	0.000	1.209	1.000	8.000	0.000	2.268

Table 5.9 Query Abandonment Descriptive Statistics

in non-Web tabs (i.e., Images, Videos, and News tabs) ($U = 200$, $p = 1.00$); (4) number of distinct SERP results clicked ($U = 155$, $p = 0.22$); nor (5) average rank of SERP results clicked ($U = 194$, $p = 0.87$). Table 5.10 shows the median, maximum, minimum, and standard deviation of click characteristics for each group.

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Clicks	7.500	28.000	1.000	6.871	10.500	20.000	2.000	4.342
Web Tab Clicks	5.500	22.000	1.000	5.139	9.000	15.000	2.000	3.703
Non-Web Tab Clicks	0.000	15.000	0.000	3.908	0.000	9.000	0.000	3.000
Distinct Clicks	7.500	25.000	1.000	6.052	10.500	17.000	2.000	3.813
Avg. Click Rank	2.481	7.933	1.000	1.738	2.929	6.300	1.000	1.320

Table 5.10 SERP Click Characteristics Descriptive Statistics

5.4.4 Pace of Interaction

Figures 5.39-5.42 show differences in pace of interaction between groups. There were no statistically significant differences detected between groups in terms of: (1) time (in seconds) between events ($U = 225$, $p = 0.51$); (2) completion time or duration (in minutes) of the search session ($U = 167$, $p = 0.38$); (3) number of queries per minute ($U = 177$, $p = 0.55$); nor (4) number of clicks per minute ($U = 180$, $p = 0.60$). Table 5.11 shows the median, maximum, minimum, and standard deviation of pace of interaction for each group.

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Time btwn. Events	114.784	265.889	34.588	70.034	99.805	320.000	50.159	63.808
Search Duration	27.408	39.883	10.15	10.491	33.483	41.233	21.10	6.436
Queries per Minute	0.269	1.071	0.076	0.234	0.306	0.652	0.121	0.159
Clicks per Minute	0.302	0.875	0.099	0.223	0.341	0.556	0.054	0.135

Table 5.11 Pace of Interaction Descriptive Statistics

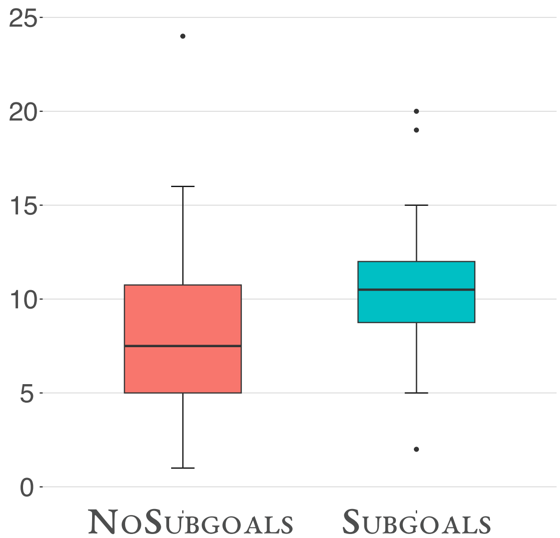

Figure 5.34 Clicks (number of SERP clicks)

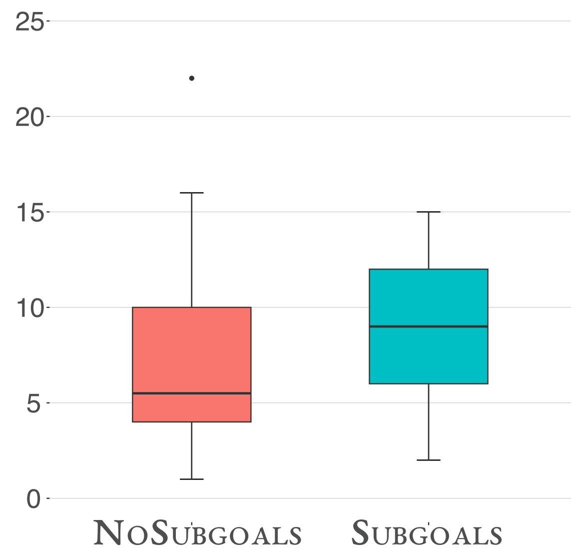

Figure 5.35 Web clicks (number of SERP clicks within the main Web tab)

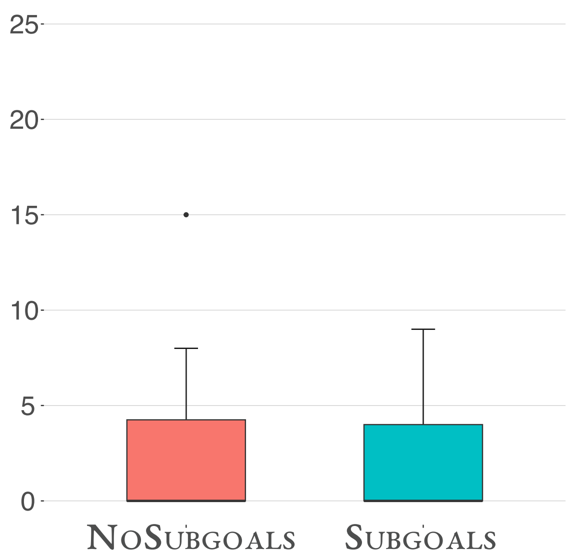

Figure 5.36 Non-web clicks (number of SERP clicks within the Images, Video, and News tabs)

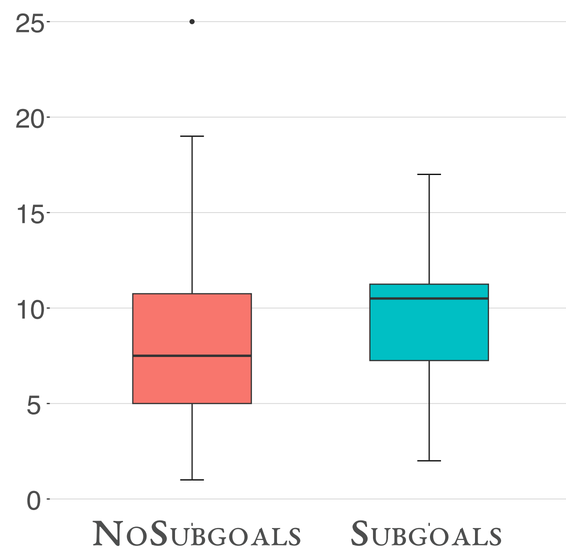

Figure 5.37 Distinct URLs (number of unique SERP results clicked by a participant)

5.4.5 Singularity of Interaction

Figures 5.43-5.45 show differences in singularity of interaction between groups. On average, there were more unique queries issued in the Subgoals versus NoSubgoals condition. There was a marginally significant difference detected between groups ($U = 128$, $p = 0.05$).

There were no statistically significant differences detected between groups in terms of: (1) unique

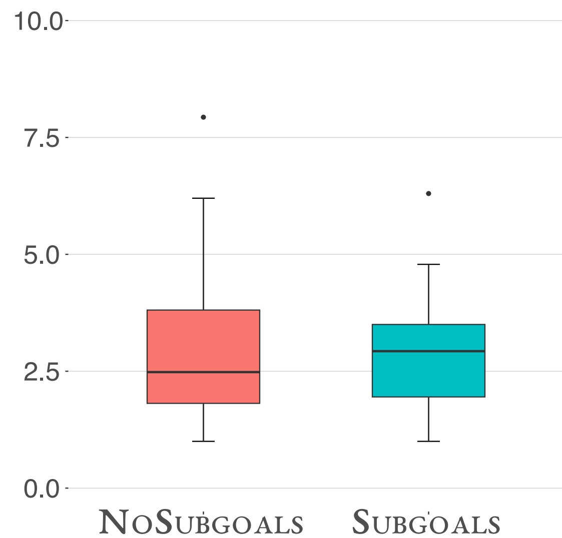

Figure 5.38 Average click rank (average rank across all search results clicked)

query terms ($U = 192$, $p = 0.84$) nor (2) unique SERP results clicked (i.e., results not clicked by any other participant) ($U = 190$, $p = 0.78$). Table 5.12 shows the median, maximum, minimum, and standard deviation of singularity of interaction for each group.

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Unique Queries*	**3.000**	**16.000**	**0.000**	**5.099**	**5.000**	**17.000**	**2.000**	**4.030**
Unique Qry. Terms	2.000	12.000	0.000	3.768	2.000	8.000	0.000	1.960
Unique URLs	4.000	16.000	0.000	4.477	4.000	8.000	0.000	2.397

Table 5.12 Singularity of Interaction Descriptive Statistics (* indicates marginal statistical significance)

5.5 Differences in SRL Processes

In this section, I present results addressing **RQ4** associated with SRL processes. In my study, SRL processes were coded from think-aloud comments and search environment behaviors in micro-SRL and macro-SRL processes. These results include the frequency of macro-SRL processes, frequency of micro-SRL processes, and diversity of macro-SRL processes (i.e., the number of unique micro-SRL processes engaged within a particular macro-SRL process).

First, I present the results of frequencies of macro-SRL processes. I provide differences between groups (i.e., NoSubgoals versus Subgoals) of frequencies of: (1) *Planning*; (2) *Strategy Use*; (3)

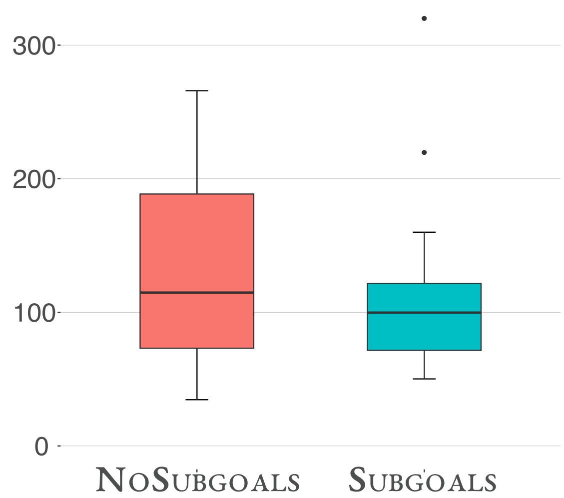

Figure 5.39 Time between events (average time in seconds between subsequent events [i.e., queries and clicks])

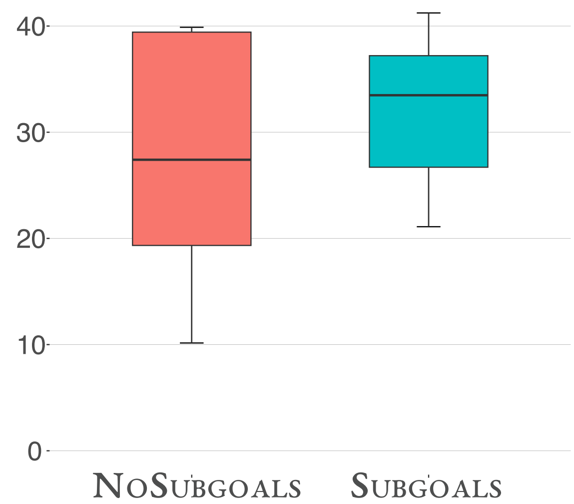

Figure 5.40 Search completion time (session length from first query to exiting search system in minutes)

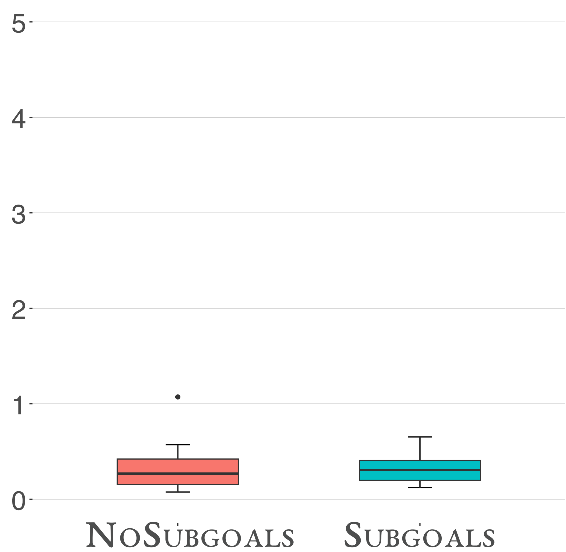

Figure 5.41 Queries per minute (rate of queries calculated from dividing total queries by search completion time)

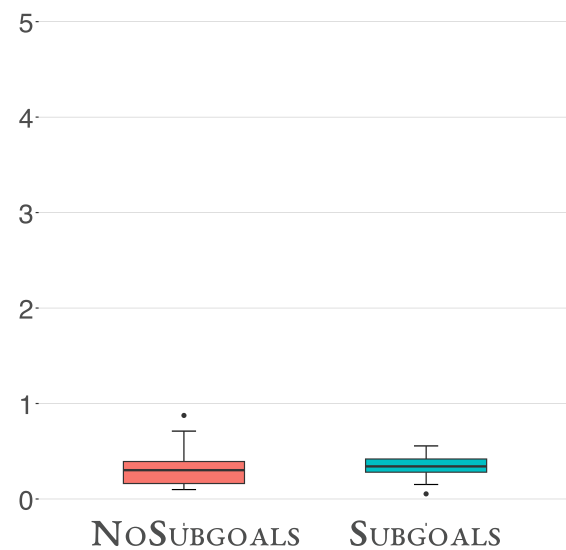

Figure 5.42 Clicks per minute (rate of clicks calculated from dividing totals clicks by search completion time)

Monitoring; and (4) *Interest*.

Second, I present the results of diversity of macro-SRL processes. I provide differences between groups (i.e., NoSubgoals versus Subgoals) of diversity of micro-SRL processes within: (1) *Planning*; (2) *Strategy Use*; and (3) *Monitoring*.

Third, I present the results of frequencies of micro-SRL processes. In this section, I only provide

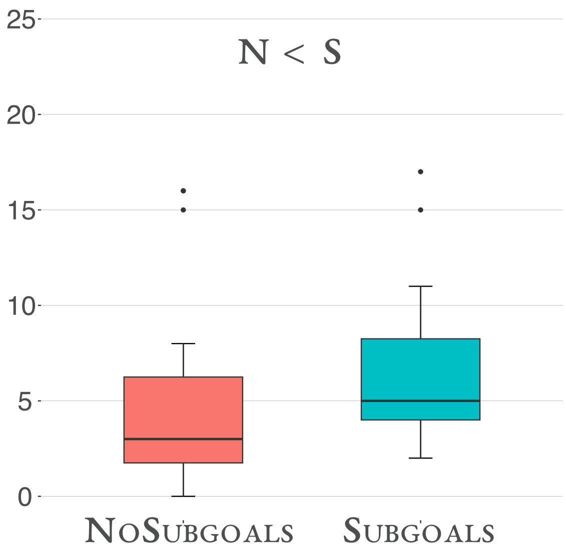

Figure 5.43 Unique queries (queries not issued by any other participant)

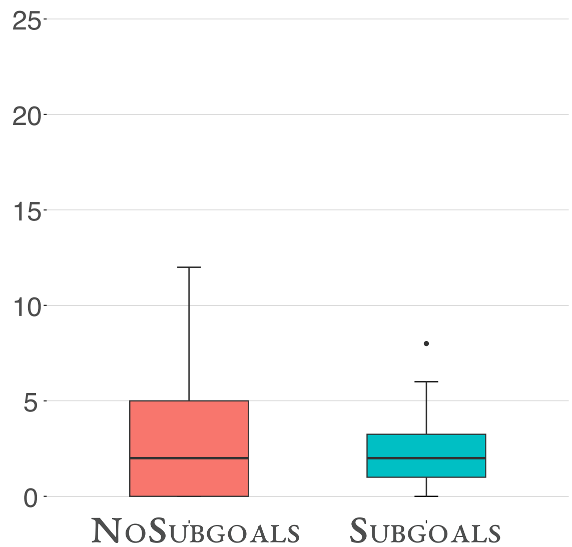

Figure 5.44 Unique query terms (query terms not issued by any other participant)

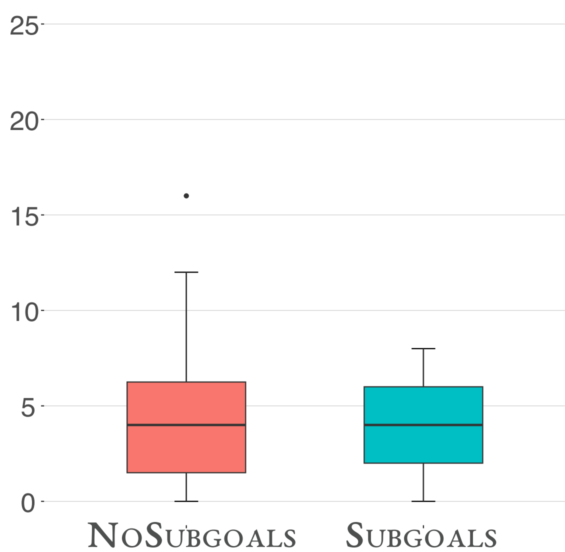

Figure 5.45 Unique URLs (number of SERP click results not clicked by any other participant)

the differences of micro-SRL processes that were significant or marginally significant. This section is included in order to help explore which specific micro-SRL processes were different within the macro-SRL processes found to be different between groups. I provide differences between groups (i.e., NoSubgoals versus Subgoals) of frequencies of: (1) *Modifies Subgoals*; (2) *Revisits Previous Subgoal*; (3) *Recycle Goal in Working Memory*; (4) *Subgoals*; (5) *Comparing & Contrasting*; (6)

Prior Knowledge Activation; (7) *Expectation of Adequacy of Content*; (8) *Monitors Progress Toward Subgoals*; (9) *Monitor Subgoal Quality*; and (10) *Time Monitoring.*

5.5.1 Frequency of Macro-SRL Processes

Figures 5.46-5.49 show differences in the frequency of macro-SRL processes between groups. On average, there were more *Planning* macro-SRL processes in the SUBGOALS versus NOSUBGOALS condition. There was a statistically significant difference detected between groups ($U = 21.5$, $p < 0.0001$). On average, there were also more *Strategy Use* macro-SRL processes in the SUBGOALS versus NOSUBGOALS condition. There was a marginally significant difference detected between groups ($U = 141.5$, $p = 0.06$). On average, there were also more *Monitoring* macro-SRL processes in the SUBGOALS versus NOSUBGOALS condition. There was a statistically significant difference detected between groups ($U = 114$, $p < 0.05$).

There was no statistically significant difference detected between groups in terms of the *Interest* macro-SRL process ($U = 179$, $p = 0.26$). Table 5.13 shows the median, maximum, minimum, and standard deviation of frequency of macro-SRL processes for each group.

Variable	NOSUBGOALS				SUBGOALS			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Planning*	**8.000**	**14.000**	**3.000**	**3.183**	**21.000**	**44.000**	**6.000**	**9.110**
Strategy Use*	**45.000**	**89.000**	**8.000**	**20.861**	**56.000**	**111.000**	**31.000**	**17.584**
Monitoring*	**20.500**	**60.000**	**4.000**	**13.479**	**28.500**	**67.000**	**13.000**	**15.813**
Interest	0.000	3.000	0.000	1.046	0.000	5.000	0.000	1.572

Table 5.13 Macro-SRL Processes Descriptive Statistics (* indicates statistical or marginal statistical significance)

5.5.2 Diversity of Micro-SRL Processes

Figures 5.50-5.52 show differences in the diversity of micro-SRL processes between groups. On average, there was a greater diversity of *Planning* micro-SRL processes in the SUBGOALS versus NOSUBGOALS condition. There was a statistically significant difference detected between groups ($U = 71$, $p < 0.001$). On average, there was also a greater diversity of *Strategy Use* micro-SRL processes in the SUBGOALS versus NOSUBGOALS condition. There was a marginally significant difference detected between groups ($U = 000$, $p = 0.06$).

There was no statistically significant difference detected between groups in terms of the diversity of *Monitoring* micro-SRL processes ($U = 162$, $p = 0.15$). Table 5.14 shows the median, maximum,

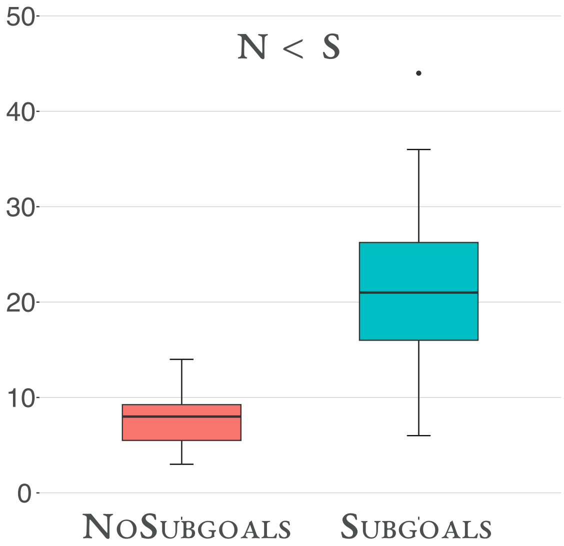

Figure 5.46 Frequency of Planning Macro-SRL processes

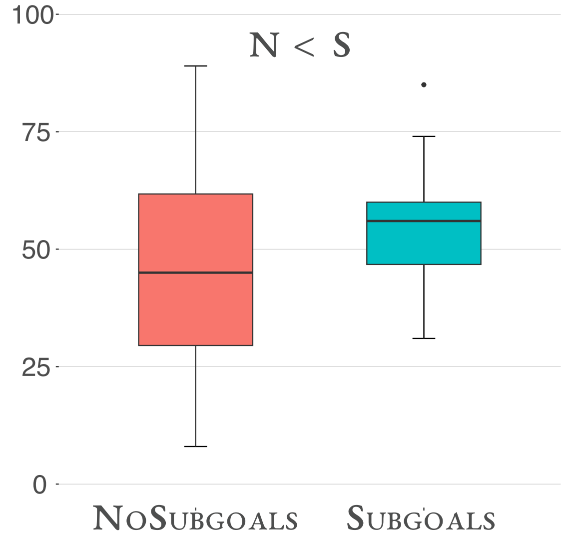

Figure 5.47 Frequency of Strategy Use Macro-SRL processes

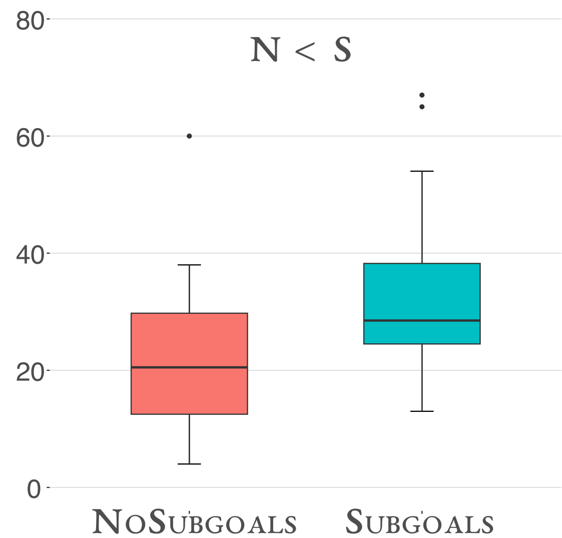

Figure 5.48 Frequency of Monitoring Macro-SRL processes

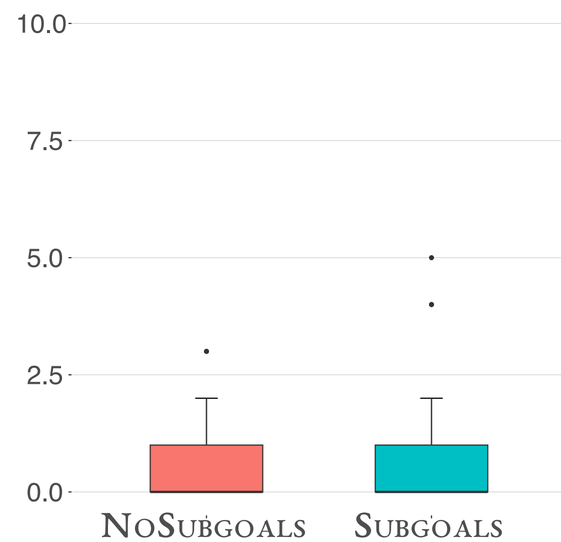

Figure 5.49 Frequency of Interest Macro-SRL processes

minimum, and standard deviation of diversity of micro-SRL processes for each group.

5.5.3 Frequency of Micro-SRL Processes

In this section, I report on the frequency of micro-SRL processes between groups. To conserve space, I only report on differences that were statistically significant or marginally significant. This analysis was conducted to better understand which micro-SRL processes may have led to differences between macro-SRL processes between groups.

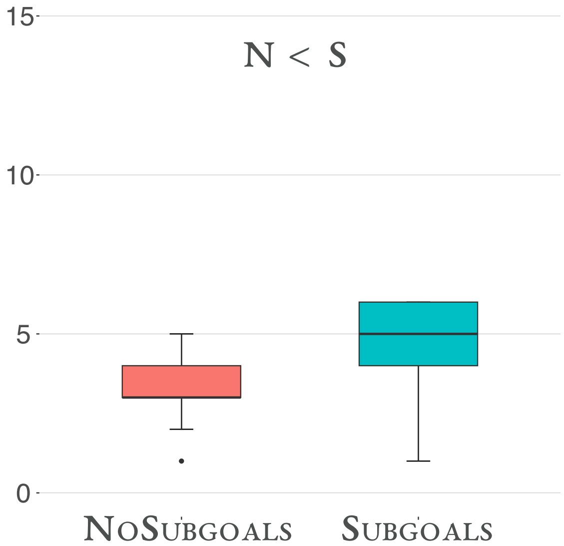

Figure 5.50 Diversity of Planning Micro-SRL Processes

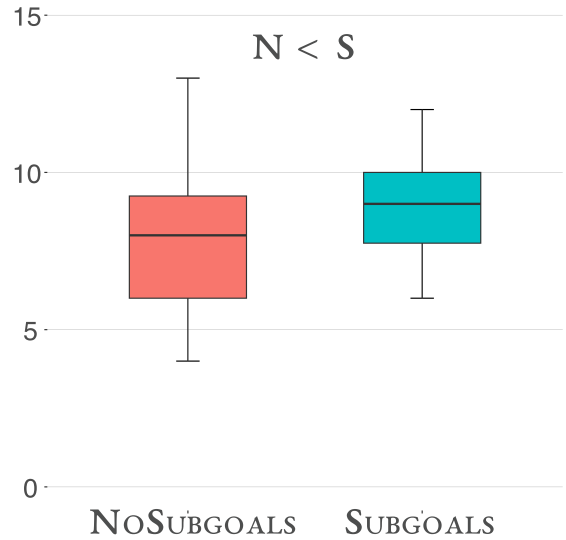

Figure 5.51 Diversity of Strategy Use Micro-SRL Processes

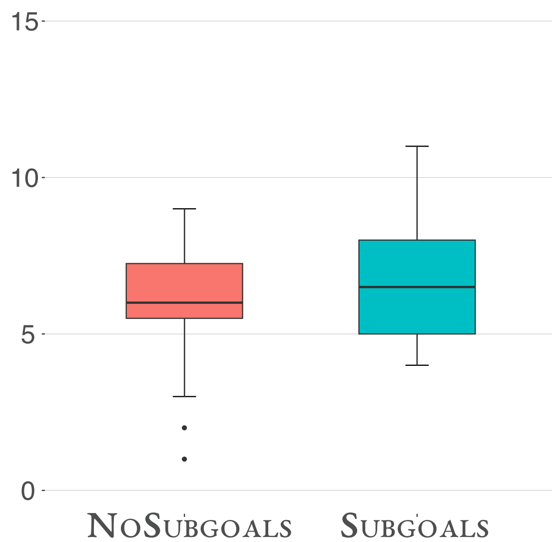

Figure 5.52 Diversity of Monitoring Micro-SRL Processes

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Planning*	**3.000**	**5.000**	**1.000**	**0.875**	**5.000**	**6.000**	**1.000**	**1.268**
Strategy Use*	**8.000**	**13.000**	**4.000**	**2.564**	**9.000**	**12.000**	**6.000**	**1.835**
Monitoring	6.000	9.000	1.000	2.167	6.500	11.000	4.000	1.765

Table 5.14 Diversity of Micro-SRL Processes Descriptive Statistics (* indicates statistical or marginal statistical significance)

Planning Micro-SRL Processes

Figures 5.53-5.56 show differences in frequency of *Planning* micro-SRL processes between groups. There *were* statistically significant differences detected between groups in terms of the frequency of: (1) *Modifies Subgoals* ($U = 67$, $p < 0.0001$); (2) *Revisits Previous Subgoal* ($U = 25$, $p < 0.0001$); (3) *Recycle Goal in Working Memory* ($U = 114$, $p < 0.01$); and (4) *Subgoals* ($U = 64$, $p < 0.001$). Table 5.15 shows the median, maximum, minimum, and standard deviation of frequency of *Planning* micro-SRL processes for each group.

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Modifies Subgoals*	**0.000**	**1.000**	**0.000**	**0.224**	**1.000**	**10.000**	**0.000**	**2.397**
Revisits Previous Subgoal*	**0.500**	**3.000**	**0.000**	**1.021**	**6.000**	**16.000**	**0.000**	**4.033**
Recycle Goal in Working Memory*	**0.000**	**3.000**	**0.000**	**0.821**	**1.000**	**2.000**	**0.000**	**0.768**
Subgoals*	**4.500**	**9.000**	**0.000**	**2.296**	**7.500**	**17.000**	**4.000**	**3.859**

Table 5.15 *Planning* Micro-SRL Processes Descriptive Statistics (* indicates statistical significance)

Strategy Use Micro-SRL Processes

Figures 5.57 and 5.58 show differences in frequency of *Strategy Use* micro-SRL processes between groups. There *were* statistically significant differences detected between groups in terms of the frequency of: (1) *Comparing & Contrasting* ($U = 142$, $p = 0.05$) and (2) *Prior Knowledge Activation* ($U = 114$, $p < 0.01$). Table 5.16 shows the median, maximum, minimum, and standard deviation of frequency of *Strategy Use* micro-SRL processes for each group.

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Comparing & Contrasting*	**0.000**	**5.000**	**0.000**	**1.663**	**1.500**	**6.000**	**0.000**	**1.725**
Prior Knowledge Activation*	**0.500**	**3.000**	**0.000**	**1.309**	**3.000**	**9.000**	**0.000**	**2.758**

Table 5.16 *Strategy Use* Micro-SRL Processes Descriptive Statistics (* indicates statistical or marginal statistical significance)

Monitoring Micro-SRL Processes

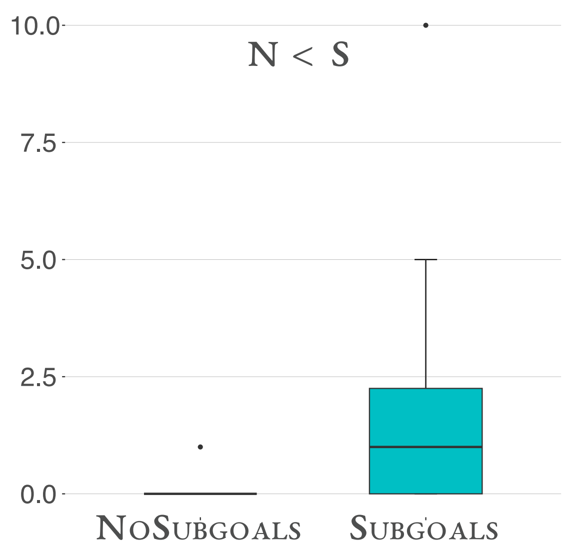

Figure 5.53 Frequency of Modifies Subgoals Micro-SRL processes

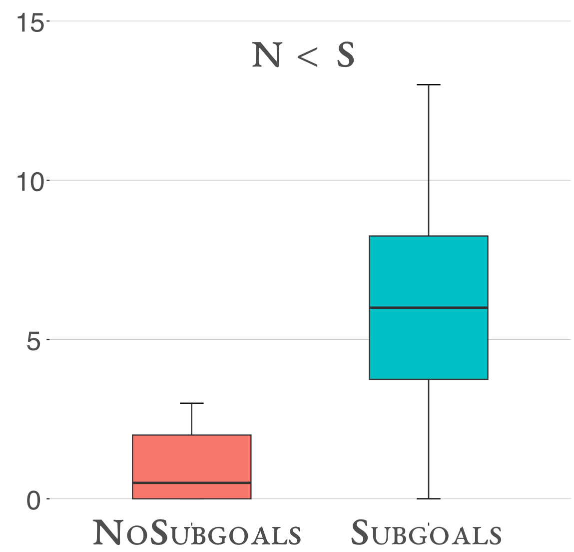

Figure 5.54 Frequency of Revisits Previous Subgoal Micro-SRL processes

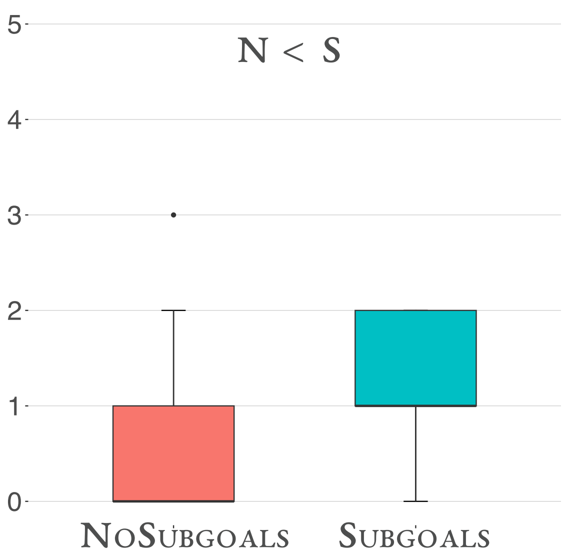

Figure 5.55 Frequency of Recycle Goal in Working Memory Micro-SRL processes

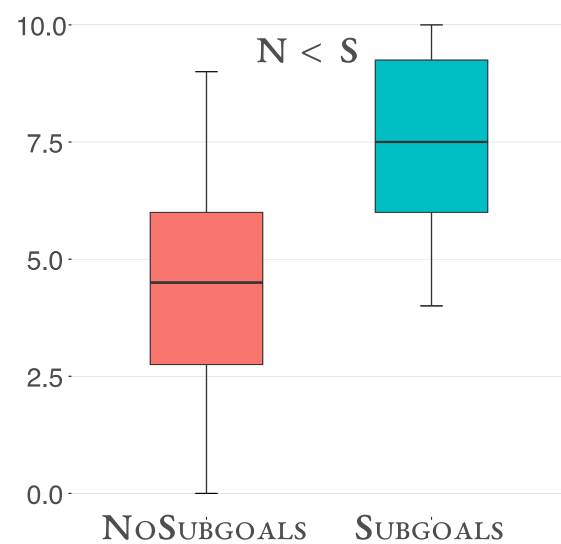

Figure 5.56 Frequency of Subgoals Micro-SRL processes

Figures 5.59–5.62 show differences in frequency of *Monitoring* micro-SRL processes between groups. There *were* statistically significant differences detected between groups in terms of the frequency of: (1) *Expectation of Adequacy of Content* ($U = 109$, $p < 0.01$); (2) *Monitor Progress Toward Subgoals* ($U = 55$, $p < 0.0001$); (3) *Monitor Subgoal Quality* ($U = 160$, $p < 0.05$); and (4) *Time Monitoring* ($U = 101$, $p < 0.001$). Table 5.17 shows the median, maximum, minimum, and standard deviation of frequency of *Monitoring* micro-SRL processes for each group.

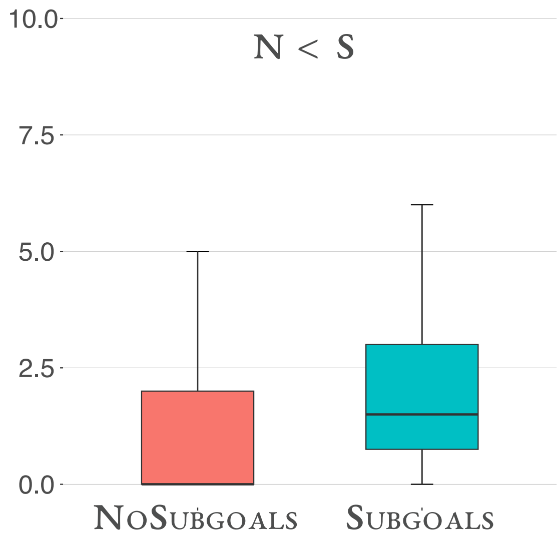

Figure 5.57 Frequency of Comparing & Contrasting Micro-SRL processes

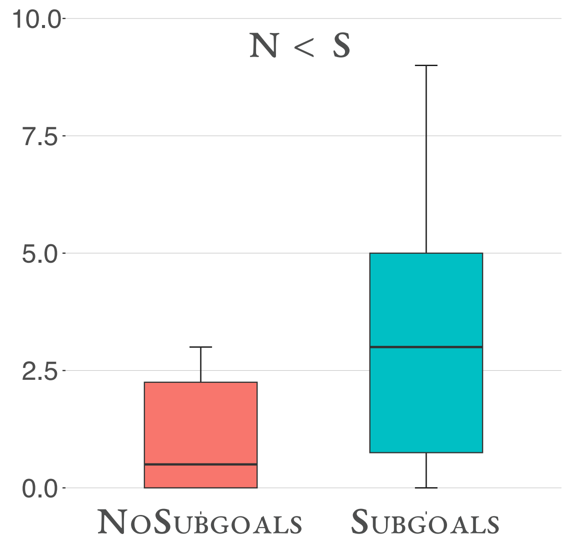

Figure 5.58 Frequency of Prior Knowledge Activation Micro-SRL processes

Variable	NoSubgoals				Subgoals			
	Median	*Max.*	*Min.*	*SD*	*Median*	*Max.*	*Min.*	*SD*
Expectation of Adequacy of Content*	**3.500**	**23.000**	**0.000**	**5.717**	**8.000**	**18.000**	**0.000**	**4.617**
Monitor Progress Toward Subgoals*	**2.000**	**3.000**	**0.000**	**1.040**	**4.500**	**15.000**	**1.000**	**3.401**
Monitor Subgoal Quality*	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**2.000**	**0.000**	**0.657**
Time Monitoring*	**0.000**	**3.000**	**0.000**	**0.696**	**1.000**	**9.000**	**0.000**	**2.134**

Table 5.17 *Monitoring* Micro-SRL Processes Descriptive Statistics (* indicates statistical significance)

5.6 Summary

Table 5.18 summarizes all significant and marginally significant results from **RQ1–RQ4**. In all cases, values were greater in the Subgoals versus NoSubgoals condition.

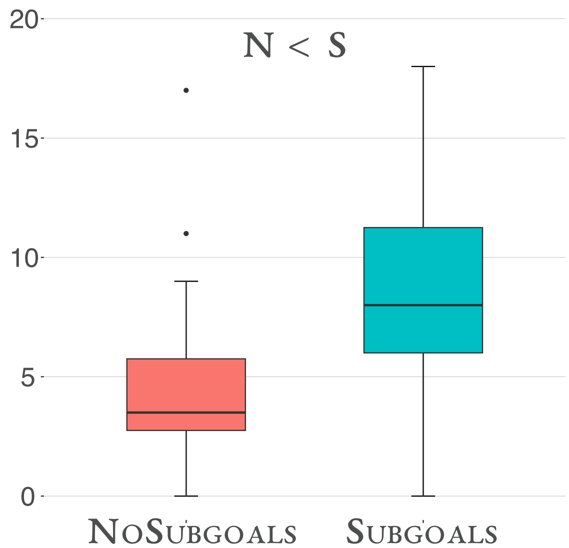

Figure 5.59 Frequency of Expectation of Adequacy of Content Micro-SRL processes

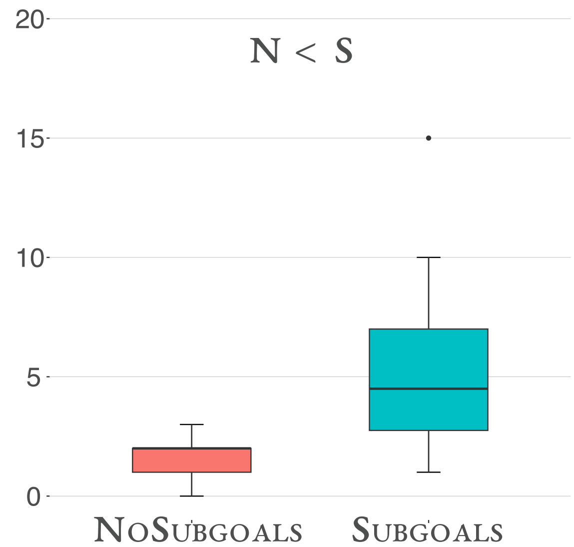

Figure 5.60 Frequency of Monitor Progress toward Subgoals Micro-SRL processes

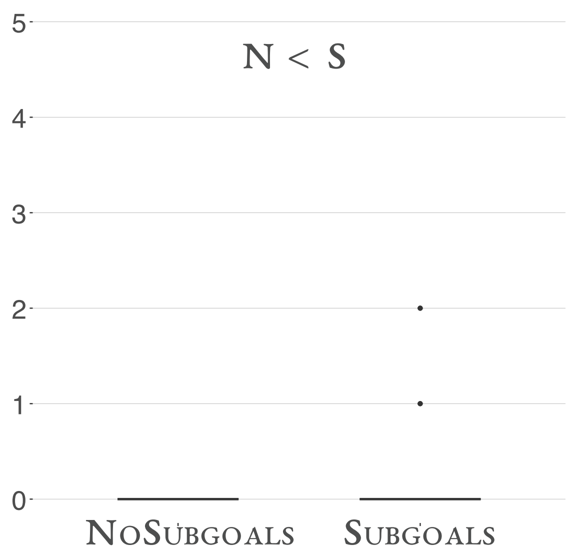

Figure 5.61 Frequency of Monitor Subgoal Quality Micro-SRL processes

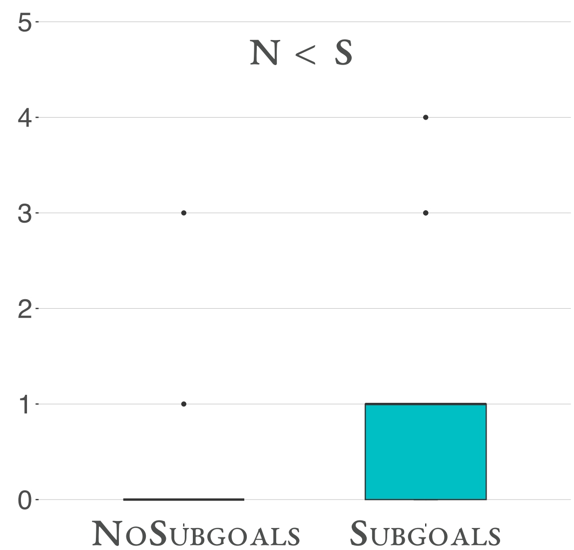

Figure 5.62 Frequency of Time Monitoring Micro-SRL processes

Table 5.18 Summary of results from **RQ1 – RQ4** (N = NoSubgoals and S = Subgoals)).

RQ	Variable Group	Variable	Result
RQ1 Learning Outcomes	ODCA	Normalized retention gain (ODCA)	N < S ($p < 0.05$)
	Open-Ended	Percent of true statements on retention (open-ended)	N < S ($p < 0.05$)
		Percent of true statements retained (open-ended)	N < S ($p < 0.001$)
RQ2 Searcher Perceptions	SRL Support	*Planning*	N < S ($p < 0.05$)
		Evaluating Progress	N < S ($p < 0.05$)
RQ3 Searcher Behaviors	Singularity of Interaction	Unique queries not issued by any other participant	N < S ($p = 0.05$)
RQ4 SRL Processes	Frequency of Macro-SRL	*Planning*	N < S ($p < 0.001$)
		Strategy Use	N < S ($p = 0.06$)
		Monitoring	N < S ($p < 0.05$)
	Diversity of SRL Processes	*Planning* micro-SRL processes	N < S ($p < 0.001$)
		Strategy Use micro-SRL processes	N < S ($p = 0.06$)
	Frequency of *Planning* Micro-SRL	*Modifies Subgoals*	N < S ($p < 0.001$)
		Revisits Previous Subgoal	N < S ($p < 0.001$)
		Recycle Goal in Working Memory	N < S ($p < 0.01$)
		Subgoals	N < S ($p < 0.001$)
	Frequency of *Strategy Use* Micro-SRL	*Comparing & Contrasting*	N < S ($p = 0.05$)
		Prior Knowledge Activation	N < S ($p < 0.01$)
	Frequency of *Monitoring* Micro-SRL	*Expectation of Adequacy of Content*	N < S ($p < 0.01$)
		Monitor Progress Toward Subgoals	N < S ($p < 0.001$)
		Monitor Subgoal Quality	N < S ($p < 0.05$)
		Time Monitoring	N < S ($p < 0.001$)

CHAPTER 6

Discussion

In my dissertation study, I investigated the influence of subgoals on learning during search across four research questions (Chapter 3). In the following sections, I consider *why* significant differences between groups may have been found in results from **RQ1**–**RQ4**. Additionally, I discuss the implications of the findings associated with **RQ1**–**RQ4**. Finally, I discuss directions for future work from my dissertation study.

6.1 RQ1 Findings & Implications

RQ1 investigated the influence of the subgoal condition on learning during search. The investigation of learning included two different types of assessments: a multiple-choice assessment called the ODCA and an open-ended assessment. These assessments were administered at multiple points in time. The ODCA was administered before, immediately after, and one week after the search session. The open-ended assessment was administered immediately after and one week after the search session. Using assessments from different points in time, **RQ1** investigated both immediate post-task learning outcomes and retention learning outcomes.

H1 stated that participants in the Subgoals condition would have higher learning outcomes. Participants in the Subgoals condition *did* have higher learning outcomes. Immediate post-task scores on the ODCA and open-ended assessments were not significantly different between groups. However, immediate post-task median scores were higher in the Subgoals condition, and had lower variance, on both the ODCA and open-ended assessments. Retention assessments were significantly different between groups on *both* the ODCA and open-ended assessments. The normalized gain on the ODCA retention assessment was higher from participants in the Subgoals condition. Additionally, participants in the Subgoals condition wrote a greater percentage of true statements on the retention assessment and retained a greater percentage of true statements from the post-task assessment to the retention assessment one week later.

Given these results, it seems that subgoals may have positively influenced participants' learning during search and, in particular, retention of learning after the search session. Goals have several important functions that support learning. First, goals prompt learners to consider their task understanding, i.e., macro-SRL *Planning.* Second, goals direct attention toward task-related activities and strategies, i.e., macro-SRL *Strategy Use.* Third, goals provide standards to monitor progress and evaluate performance toward the overall task goal, i.e., macro-SRL *Monitoring.* Participants in the SUBGOALS condition perceived greater support for SRL. Perhaps greater support for SRL increased learning retention. This resonates with prior work that has shown supporting SRL increases outcomes on delayed retention tests [146]. Additionally, participants in the SUBGOALS condition had higher frequencies and a greater diversity of SRL processes based on their think-aloud comments and behaviors. Perhaps the higher rate of SRL processes and greater diversity of SRL processes enabled participants in the SUBGOALS condition to better retain what was learned during the search session. In particular, more frequent engagement in *Monitoring* may have indirectly improved learning outcomes. In prior work, learners that more frequently engaged in *Monitoring* processes also used more deep-level learning strategies (e.g., *Knowledge Elaboration*) and had greater conceptual knowledge gains [147, 148]. Perhaps, similarly, participants in the SUBGOALS condition had higher learning outcomes because they more frequently engaged in *Monitoring* processes and more deep-level strategies, possibly evidenced by their greater diversity of *Strategy Use* micro-SRL processes.

Important implications stem from **RQ1**. First, results indicate that the development and implementation of subgoals seems to increase frequency of macro-SRL processes (e.g., increases *Planning* macro-SRL processes during subgoal development, increasing *Monitoring* macro-SRL processes when monitoring subgoal progress) which may, subsequently, increase learning retention. Search systems should support searchers using tools like the Subgoal Manager in developing, following, and achieving subgoals to assist in learning during search.

Second, although statistically significant differences were not found on immediate post-task assessments, significant differences were found in *both* retention assessments. This result is echoed by prior work that has found no significant differences in immediate post-task assessments, but *have* found significant differences in delayed retention assessments [149]. Research in search-as-learning have administered many immediate post-tests, but very few delayed retention tests. My study highlights the potential importance of capturing retention of learning.

Finally, while not specifically analyzed in the dissertation, the use of more than one assessment (i.e., ODCA and open-ended) captures a greater breadth of what was learned during search. While the ODCA provided a simple mechanism for directly comparing groups, the open-ended assessment provided insight into the variety of what was learned across groups. Overall, 374 unique statements were identified across participants. Further analysis could identify the variation of topics and complexity of statements to better understand potential differences in learning between groups. Additionally, learning retention scores were higher in the Subgoals condition and significantly different between groups. In this way, the use of more than one assessment provided stronger evidence of increased learning retention outcomes than if only one assessment had been administered. Researchers in search-as-learning may also strengthen their findings in similar ways (i.e., greater breadth of learning captured, additional evidence of learning) through implementing more than one learning assessment.

6.2 RQ2 Findings & Implications

RQ2 investigated the influence of the subgoal condition on searcher perceptions of interest increase, knowledge increase, satisfaction, difficulty, SRL support, and engagement. Searcher perceptions were collected through a post-task questionnaire immediately after the search session.

H2 stated that participants in the Subgoals condition will report higher levels of SRL support. Participants in the Subgoals condition did, indeed, report higher levels of SRL support in terms of *Planning* and *Evaluating Progress* (a subset of *Monitoring* SRL processes). Participants in the Subgoals condition were asked to develop subgoals and were provided with the Subgoal Manager to write subgoals. This may have caused participants to perceive greater support for *Planning*. Additionally, the Subgoal Manager provided text editors associated with subgoals to track how much information had been saved with respect to each subgoal and a "Subgoal Complete" checkbox to prompt evaluation of progress toward subgoal achievement. These additional features may have caused participants to perceive greater support for *Evaluating Progress*. These results resonate with my preliminary study A in which participants in the SelfSetSubgoals condition reported greater support for *Planning* and *Evaluating Progress* than participants in the NoSubgoals condition.

Interestingly, perceptions of *Monitoring* were not significantly different between groups. However, *actual* frequencies of macro-SRL *Monitoring* processes were higher in the Subgoals versus NoSubgoals condition. Prior work has questioned the reliability of self-report SRL items from questionnaires as a *proxy* for actual SRL. These results highlight the importance of coding think-aloud comments and behaviors when researchers are aiming to better understand *actual observed* SRL processes.

Searcher perceptions of interest increase, knowledge increase, satisfaction, difficulty, and engagement were exploratory and did not have specific directional hypotheses. None of these perception variables were significantly different between groups. The dissertation study was a between subjects design, meaning each participant was only exposed to a single condition. If participants were exposed to both conditions, it would have been more likely to see differences across more dimensions of searcher perceptions (i.e., participants' perceptions for the second task would have been grounded in their perceptions' of the first task). Additionally, participants in the NoSubgoals condition were allowed to approach the search session in whatever way they were comfortable. Differences in difficulty and perceived usability (a measure of engagement) were detected in my preliminary study A with participants in the AssignedSubgoals condition reporting greater difficulty and lower perceived usability. However, there were no significant differences of these perceptions between SelfSetSubgoals and NoSubgoals. Perhaps the relative freedom in both conditions of the current study resulted in low rates of difficulty and relatively high rates of perceived usability.

6.3 RQ3 Findings & Implications

RQ3 investigated the influence of the subgoal condition on search behaviors. Search behaviors were captured on the SERP along five dimensions: query characteristics; query abandonment; SERP click characteristics; pace of interaction; and singularity of interaction. All search behavior variables were exploratory and did not have specific directional hypotheses.

Of all 20 variables within the 5 dimensions, only 1 variable was significantly different between groups. Participants in the Subgoals condition issued more unique queries (i.e., queries not issued by any other participant). Participants in the Subgoals condition developed specific subgoals (i.e., subgoals with specific content, action, standard, and time). This level of specificity may had led participants in the Subgoals condition to have more specific queries unique to their particular learning subgoals. This suggests that search systems should help searchers develop specific, learning-oriented goals *and* formulate specific queries that address particular subgoals to support learning during search. Potentially, a search system could use content and standards set in subgoals to help a searcher formulate a more unique, individualized query for each subgoal.

No significant differences were detected between groups of the remaining 19 search behavior variables. Although learning was different between groups, there are potential explanations as to why search behavior differences were not detected. First, the extent to which search behaviors predict learning may depend on the task. Studies that have found search behaviors to predict learning have asked participants to complete relatively simple, factual knowledge tasks [150, 24, 26]. In this study, participants completed a complex, conceptual knowledge task. Our task required participants to engage in complex cognitive processes, such as browsing for information, reading, note-taking, reflecting, etc. Such complex cognitive processes may not be associated with behaviors that can be captured by the search system (e.g., queries and clicks). Second, this study consisted of a small sample size. A larger sample size may highlight differences that are more difficult to detect in smaller samples.

6.4 RQ4 Findings & Implications

RQ4 investigated the influence of the subgoal condition on macro-SRL and micro-SRL processes. Participant behaviors and think-aloud comments captured in Zoom recordings were coded into micro-SRL (and macro-SRL) processes.

H4 stated participants would engage SRL processes more frequently in the Subgoals condition. Frequencies of 3 macro-SRL and 10 micro-SRL processes were significantly different or marginally different between groups (Table 5.18). Participants in the Subgoals condition had higher frequencies of these macro-SRL and micro-SRL processes.

Participants in the Subgoals condition engaged in more *Planning* macro-SRL processes. Within *Planning*, participants in the Subgoals condition engaged in more *Modifies Subgoals*, *Revisits Previous Subogoal*, *Subgoals*, and *Recycle Goal in Working Memory*. Participants in the Subgoals condition were asked to develop at least three subgoals. Some of the differences of *Planning* between groups may be accounted for by this fact. It is important to note, however, that participants in the NoSubgoals condition *did* develop subgoals, just at a lower rate than participants in the Subgoals condition (NoSubgoals $Med. = 4.5$ and Subgoals $Med. = 7.5$). This is important because it indicates that while subgoals are beneficial to searchers, they may not naturally develop them. Search systems could support searchers by prompting them to develop and modify specific subgoals across a search session.

Interestingly, *Recycle Goal in Working Memory* (*RGWM*) was different between groups. Although not specifically targeted in the Subgoals condition, participants in this condition had higher rates of *RGWM* micro-SRL process. There are two potential explanations for this difference. First, participants in the Subgoals condition may have been more likely to *have* something to recycle. In other words, participants that developed subgoals may have had clearer, more defined aims that were able to be articulated verbally. Second, participants in the Subgoals condition used the Subgoal Manager which kept their subgoals visible at all times. Being able to see the subgoals may have prompted participants to refocus and restate their objective. Search environments should support subgoal development and provide reminders of subgoal aims. This could be accomplished by highlighting the subgoal in which a searcher is currently working or interactively inserting the subgoal above the search box in the SERP.

Participants in the Subgoals condition engaged in more *Strategy Use* macro-SRL processes. Within *Strategy Use*, participants in the Subgoals condition engaged in more *Comparing & Contrasting* and *Prior Knowledge Activation*. Participants in the Subgoals condition more frequently moved back to previous subgoals and compared their notes with new information sources. For example, while reading a video transcript about diffusion, a participant in the Subgoals condition encountered information that seemed similar to their previously stated definition of osmosis. This prompted the participant to return to their subgoal about the definition of osmosis and compare and contrast definitions on the web page with those in their notes. This indicates that subgoals may have enabled participants to both compartmentalize information more efficiently and more clearly define the boundaries of one subgoal from another using specific content and standards. These more defined boundaries may have enabled participants to realize when information may be related to one subgoal versus another (e.g., specific content related to components of osmosis versus diffusion). Search environments should support *Comparing & Contrasting* by encouraging searchers to revisit previous subgoals that may be related to current web page content. Perhaps search environments could highlight information in subgoal notes that have a high degree of semantic similarity to current web page content.

Participants in the Subgoals condition also engaged in a higher frequency of *Prior Knowledge Activation* (*PKA*). This is quite an interesting outcome given that *PKA* was not *directly* supported in the Subgoals condition. That is, participants were not explicitly asked to reflect on their prior knowledge and none of the features in the Subgoal Manager even mentioned prior knowledge. It is possible that the act of developing subgoals prompted *PKA*. This resonates with prior work in SRL and goal-setting that underscores the importance of goals to instigate task-relevant knowledge [2, 39]. Additionally, subgoals may have better contextualized search sessions to individual participants in the SUBGOALS condition. Throughout the search session, participants in the SUBGOALS condition repeatedly viewed goals they themselves had written. This may have prompted searchers to more frequently be self-reflective, including reflective of what they already knew.

Participants in the SUBGOALS condition engaged in more *Monitoring* macro-SRL processes. Within, *Monitoring*, participants in the SUBGOALS condition engaged in more *Expectation of Adequacy of Content*, *Monitor Progress Toward Subgoals*, *Monitor Subgoal Quality* and *Time Monitoring*. It is somewhat obvious why micro-SRL processes of *Monitor Progress Toward Subgoals* and *Monitor Subgoal Quality* were higher in the SUBGOALS condition as participants were directly supported in monitoring their subgoals (i.e., "Subgoal Complete" checkbox) and were asked to develop subgoals with particular characteristics. This is an important result, however, in that those mechanisms seemed to be successful in increasing the behaviors of evaluating progress toward subgoals and developing high quality subgoals. It is not so intuitive, however, to understand why *Expectation of Adequacy of Content* (*EAC*) and *Time Monitoring* were more frequent of participants in the SUBGOALS condition. Participants in the SUBGOALS condition had objectives with clearly defined content and standards. In this way, clearly defined subgoals may have prompted and enabled participants to more effectively pick out information sources that were pertinent to their needs. To support *EAC*, search environments could facilitate gauging the usefulness of content by mapping it to particular learning-oriented subgoals. The search environment could always consider the words in a particular subgoal regardless of what a searcher queries.

Participants in the SUBGOALS condition also more frequently engaged in *Time Monitoring*. There may be a couple of explanations for this phenomenon. First, participants were asked to develop subgoals with a specific timeframe. Referencing subgoals across the search session that contained a specific timeframe may have caused participants to more frequently monitor time. Second, even if a participant had not set a specific time, subgoals may have increased *Time Monitoring* because participants had set a particular amount of subgoals. Having a set amount of subgoals also serves the purpose of partitioning time. As participants completed subgoals, time still had to be allotted for the remaining subgoals. This result underscores the importance of incorporating subgoals into search environments, even if simplistically implemented (e.g., a basic list alongside a SERP).

While not specifically stated as part of **RQ4**, I also investigated the diversity of micro-SRL processes within each macro-SRL process. This was done to capture differences observed during SRL coding in the variety of micro-SRL processes engaged. Participants in the Subgoals condition engaged in a greater variety of micro-SRL processes within *Planning* and *Strategy Use*. Participants in the Subgoals condition were asked to develop subgoals with specific standards. It is possible that standards prompted participants to enact more effective and different strategies in order to meet standards. For example, participants in the Subgoals condition often added a specific number of examples or differences/similarities they wanted to find relating to osmosis and diffusion. This specific number sometimes prompted them to engage *Reading Notes* to see if they had met the goal or take inventory of what they already knew about a topic, engaging *Prior Knowledge Activation*, to reach the target number set in the subgoal.

Perhaps the greater diversity of micro-SRL processes engaged contributed to better learning outcomes. This is an important question for future work. In the next section, I explore this and other important future research directions.

6.5 Future Work

Given the results from my dissertation study, there are several important directions for future work. In this section, I discuss three avenues of future research.

6.5.1 SRL Processes and Learning Outcomes

In my dissertation, **RQ1** investigated the impact of subgoals on learning during search and **RQ4** investigated the impact of subgoals on SRL processes during search. A natural next question from my dissertation results is—what impact do SRL processes have on learning during search? In a subsequent analysis, I will investigate the types of micro-SRL processes (and macro-SRL processes) that correlated with learning gains in the immediate post-task assessments and retention assessments regardless of subgoal condition. Prior work in SRL has categorized SRL strategies by complexity, grouping into surface-level (e.g., re-reading) versus deep-level (e.g., knowledge elaboration) strategies [147]. Learners that used a higher frequency of deep-level strategies had better learning outcomes. Perhaps a similar result will emerge from my dissertation study, with more complex strategies (e.g., *Forming New Conclusion*) correlating with learning outcomes. To my knowledge, this is the first *search study* to code SRL processes from think-aloud comments and interaction data. Results from this analysis would be important to the field of search-as-learning to better understand which SRL processes may contribute to learning during search. Such findings will help to inform search environment design to support specific SRL processes to increase learning during search.

6.5.2 SRL Processes and Phases of Search Session

In my dissertation, I coded which think-aloud comments and search interactions mapped to particular SRL processes, but I also captured *when* these processes occurred during search. An important next question is—when are specific micro-SRL and macro-SRL processes likely to occur during search? Perhaps *Planning* processes were more frequent at the beginning of the search session, while *Monitoring* were more frequent toward the end? Within macro-SRL processes, perhaps particular micro-SRL processes were more frequent at different stages of search. For example, within the macro-SRL process of *Strategy Use*, it is possible that *Taking Notes* was more frequent in the beginning and middle of the search session, while *Forming New Conclusion* was more frequent toward the end of the search session. Investigating *when* SRL processes are more likely to occur will help inform when particular SRL processes should be supported for better learning during search. An additional related question is—are there particular SRL process sequences that correlate with better learning outcomes? Prior work has shown that more frequent *Monitoring* can result in better learning outcomes [147]. Perhaps those that use *Monitoring* throughout the search session have better learning outcomes than those who only engage *Monitoring* toward the end of the search session.

6.5.3 Supporting SRL in Search

Findings from my dissertation study also inform future directions of exploration for supporting SRL with the search environment. In my dissertation study, a tool called the Subgoal Manager was provided to enable participants to develop and track subgoals as well as take notes and save information associated with particular subgoals. Although the Subgoal Manager was a simple tool, it was effective in its purpose, to promote: goal-setting (e.g., *Subgoals*); subgoal-related strategies (e.g., *Note Taking* associated with subgoals); and subgoal monitoring (e.g., *Monitor Progress Toward Subgoals*). The Subgoal Manager, however, could be further developed to enrich subgoal development and subgoal progress to continue to increase learning outcomes.

In the dissertation study, I observed several ways in which participants developed subgoals that made the overall task difficult or were altogether counterproductive. First, some participants ordered subgoals in a way that was problematic to the domain of the task. The task asked that participants learn about the concepts of diffusion and osmosis. Osmosis is a specific type of diffusion. Those participants who learned about osmosis first, struggled to grasp the concept because diffusion is a sort of prerequisite knowledge for osmosis. Second, some participants developed subgoals that were not logically distinct. For example, one participant created the following subgoals: "Spend 15 minutes summarizing the specific conceptual expectations for learning Diffusion and Osmosis pertaining to high school biology in no more than 5 statements." and "Spend 20 minutes gathering the high school biology expectations for understanding Diffusion and Osmosis by compiling the information into a one-page summary." These subgoals are not logically distinct, as they both essentially ask the searcher to understand what a high school biology student should know about diffusion and osmosis. Third, some participants did not create subgoals that parallelled each other or created subgoals that skipped steps. For example, a participant created a subgoal to understand the definition of diffusion, then another subgoal to find 3 examples of diffusion, then another to understand the definition of osmosis, but did *not* create the parallel subgoal of finding 3 examples of osmosis.

In order to address these three potential issues (i.e., subgoals that are out of order, subgoals that are not logically distinct, and subgoals that skip steps) an updated Subgoal Manager could investigate the *progression* of subgoals. The Subgoal Manager could view subgoals as a set and provide feedback to the searcher as to: (1) a logical order of the subgoals (e.g., consider prerequisite knowledge); (2) whether or not subgoals are logically distinct (e.g., indicate overlap in subgoal content with highlighting); and (3) whether subgoals are in parallel or are skipping steps. This feedback could be provided interactively as subgoals are developed or as a review process once searchers have indicated they are done creating subgoals.

Finally, participants created subgoals that were either too broad (e.g., subgoal stating "Basics of diffusion") or too specific (e.g., specifying that examples of diffusion needed to relate to biology). Similar to prior work [151], the Subgoal Manager could function as a pedagogical agent and incorporate feedback after each subgoal is developed to indicate if a subgoal is too broad or too specific. The pedagogical agent could offer ways to further specify (e.g., "Instead of 'Basics of diffusion', perhaps try starting with understanding the definition of diffusion in your own words.") or broaden a particular subgoal (e.g., "You may not need examples to relate to biology in order to get an initial understanding of diffusion."). The study could implement a Wizard of Oz study design and have real learning experts act as pedagogical chat agents behind the scenes.

After investigating which micro-SRL processes correlate with learning outcomes, future research should explore ways to support these processes during search. For example, *Prior Knowledge Activation* (*PKA*) was higher in the SUBGOALS condition *and* participants in the SUBGOALS condition had higher learning outcomes. If *PKA* is found to be correlated with learning outcomes, it would be important to develop search environments that encourage *PKA*. Perhaps a search environment that offers a simple, explicit prompt of "What do you already know about this concept?" after each new query topic might encourage *PKA*. Additionally, *Comparing & Contrasting* was engaged more frequently by participants in the SUBGOALS condition. To better support *Comparing & Contrasting*, a search environment could highlight portions of text on a web page that has a certain degree of semantic similarity with particular subgoal notes to encourage searchers to clarify their current understanding of a concept using new information.

Search environments aiming to support SRL should also account for individual searcher characteristics. Greene et al. [76] notes that particular micro-SRL processes that support learning are particular to individual searchers based on characteristics such as prior knowledge. For example, a searcher with less prior knowledge may need to spend more time with *Memorization* and *Taking Notes* to acquire relevant factual knowledge. Conversely, a searcher with more prior knowledge may need to spend more time with *Comparing & Contrasting* and *Forming New Conclusions* to elaborate upon their existing knowledge. Search environments could provide support for specific micro-SRL processes based on a searcher's level of prior knowledge.

CHAPTER 7

Conclusion

This dissertation aimed to understand the impact of subgoals on learning during search. In particular, I investigated the influence of subgoals on immediate learning, learning retention, searcher perceptions, search behaviors, and SRL processes during search. Insights from this dissertation offer important implications for, and contributions to, search-as-learning research as well as multiple directions for future work in support of learning during search.

7.1 Methodological Contributions

This dissertation provides several methodological contributions. First, the dissertation demonstrates the importance of multiple learning assessments in search-as-learning research. Administering multiple learning assessments provided a more nuanced picture of what was learned (i.e., open-ended assessment), while also providing an easy method for comparing groups (i.e., multiple-choice ODCA). Using the ODCA afforded several benefits. First, the ODCA was an already established, validated instrument and required no pre-study development. Second, the ODCA was easy to administer as participants completed the assessment in about seven minutes. Finally, the ODCA was straightforward to score and allowed us to easily compare between groups (i.e., participant A got question 1 correct, participant B got question 1 wrong, etc.). However, a major drawback of the ODCA, like any other multiple-choice test, is that it focuses on specific concepts and phenomena and therefore cannot capture everything that was learned. While the open-ended assessment took slightly longer to administer and was quite intensive to score, it provided a broader picture of what each individual participant learned and retained over time.

Another important benefit of implementing multiple assessments was that they provided stronger evidence of differences in learning between groups. Both learning retention assessment scores (on ODCA and open-ended) were significantly different between groups. Participants in the Subgoals condition scored higher on *both* the ODCA and open-ended retention assessments. Having evidence of higher learning outcomes from two different assessments further backs the argument that subgoals positively influenced learning during search.

Second, the dissertation demonstrates the importance of capturing learning retention. In the dissertation study, participants completed both learning assessments immediately after the search task *and* one week later. While learning assessment scores were not significantly different on the immediate post-task assessments, they were significantly different on *both* retention assessments. Learning retention has rarely been included in prior search-as-learning work, however these results indicate that capturing learning retention is important to better understanding learning during search.

Finally, the dissertation is the first search study to code SRL processes that participants engaged with during the search session. SRL is important to improving learning outcomes [47, 48, 49, 50, 147, 148], however capturing observed SRL processes has not yet been explored in search-as-learning. To my knowledge, this is the first search study to observe and code actual SRL processes during a search session. Because of the importance of SRL and learning, I assert that coding of SRL processes should be a central focus of search-as-learning work. This dissertation offers future researchers actual examples of SRL processes to facilitate SRL coding of learning during search. The SRL processes coded in my dissertation are motivated by prior work [79, 37]. However, as part of my dissertation, I have created new micro-SRL processes that can be used in future work. Examples include: *Revisits Task*, *Questioning Task Expectation*, *Monitor Subgoal Quality*, *Revisits Previous Subgoal*, and *Modifies Subgoal*.

7.2 Technical Contributions

Several important technical contributions are provided from this dissertation. First, this dissertation shows that subgoals seem to support learning during search. Participants in the SUBGOALS condition were trained on good subgoal characteristics and provided with the Subgoal Manager, a tool that supported the development of subgoals and monitoring of subgoal progress. Participants in the SUBGOALS condition had higher learning outcomes and retained more knowledge one week after the search session. Subgoals support learning in several ways, prompting learners to: reflect on what they already know about a task; select and use task-relevant strategies; and provide standards for monitoring subgoal progress.

Second, this dissertation shows that subgoals seem to increase the rate of SRL processes during search. Participants in the SUBGOALS condition engaged in higher frequencies of macro-SRL processes of *Planning*, *Strategy Use*, and *Monitoring*. There were also many micro-SRL processes within macro-SRL processes that were significantly different between groups and more frequent by participants in the SUBGOALS condition. While some of these micro-SRL processes were explicitly supported in the SUBGOALS condition (e.g., *Subgoals*, *Revisit Previous Subgoal*, *Monitoring Subgoal Progress*) others were *not* explicitly supported. For example, participants in the SUBGOALS condition engaged in higher frequencies of *Prior Knowledge Activation*. While not explicitly supported in the SUBGOALS condition, *Prior Knowledge Activation* may have been higher because participants were inadvertently prompted to self-reflect during subgoal development and throughout the search session as subgoals were displayed in the Subgoal Manager.

Finally, this dissertation shows that subgoals seem to increase the diversity of SRL processes engaged during search. Participants in the SUBGOALS condition engaged in greater variety of micro-SRL processes within three of the four macro-SRL processes. *Planning* micro-SRL processes were explicitly supported in the SUBGOALS condition and seemed to successfully encourage a wider variety of micro-SRL processes to be engaged. Interestingly, no *Strategy Use* micro-SRL process was *uniquely* supported by the Subgoal Manager in the SUBGOAL condition. That is, micro-SRL processes supported by the Subgoal Manager in the SUBGOALS condition (e.g., *Note Taking*, *Copy Notes*) were *also* supported by the Text Editor in the NOSUBGOALS condition. In spite of this, participants engaged a more diverse set of *Strategy Use* micro-SRL processes in the SUBGOAL condition. This may have been influenced by the standards participants set in their subgoals. Standards may have prompted participants to enact a broader range of strategies in order to meet the standards.

7.3 Future Directions

There are several important directions for future work that stem from this dissertation. First, in future work, I will explore the impact of SRL processes on learning during search. This dissertation found that subgoals seemed to support both learning during search *and* SRL processes during search. My next analysis will be—what impact do SRL processes have on learning during search? Exploring this question will help inform search environments in support of particular SRL processes to promote learning during search.

Second, prior work has aimed to characterize the different phases of a search task [152, 46]. In future work, I will investigate whether specific SRL processes are more or less common during specific phases of the search task. One analysis in this dissertation involved coding of SRL processes which included data on *when* SRL processes occurred during search sessions. Using this data, I will investigate when specific micro-SRL and macro-SRL processes are likely to occur during search. Results from this exploration can inform *when* search systems should support particular SRL processes during a search session.

Finally, I will explore additional search environment features that can enhance subgoal development and learning during search. In my dissertation study, the Subgoal Manager was effective at promoting goal-setting, subgoal-related strategies, and subgoal monitoring. However, the Subgoal Manager could be further developed to continue increasing learning outcomes. For example, some participants in the SUBGOALS condition developed subgoals that made progress toward the overall task difficult. These subgoals had one or more of the following issues: (1) subgoals were not in a logical order; (2) subgoals were not logically distinct; or (3) subgoal were not in parallel or skipped steps. A future version of the Subgoal Manager could consider subgoals as a progression and offer feedback on subgoals as a *set* along these dimensions. More broadly, I will explore how search environments can support SRL processes that are particularly important for learning during search. Additionally, I will explore whether certain SRL processes are more common or important for individuals with certain characteristics.

APPENDIX A

Preliminary Study

I conducted a preliminary study to explore the role of subgoals on two main factors—(1) learning assessment scores; and (2) participant perceptions (i.e., difficulty, subgoal characteristics, SRL, and engagement during search). In particular, the study consisted of three experimental conditions AssignedSubgoals, SelfSetSubgoals, and NoSubgoals. Briefly, in the AssignedSubgoals condition, participants were provided with a version of the Subgoal Manager with pre-populated subgoals. In the SelfSetSubgoals condition, participants were provided with a version of the Subgoal Manager with blank subgoals and participants were asked to develop at least 3 subgoals with ideal subgoal characteristics. In the NoSubgoals condition, participants were provided with a plain text editor to take notes while searching and participants were not instructed to develop subgoals.

The study was conducted on Amazon Mechanical Turk in order to collect a larger sample size from each experimental condition (40 participants per condition, 120 participants total) than will be possible in the subsequent lab study. A larger sample size affords greater power and, subsequently, a higher likelihood of uncovering statistically significant differences between groups in learning assessment scores. For this reason, the preliminary study was an important first step to understanding how best to execute my dissertation study, with particular focus on the learning assessment component.

The next section provides a *brief* overview of the study protocol and associated materials (e.g., questionnaires, learning assessment, instructional videos) used in the preliminary study. Many of the materials are identical to those that will be implemented in my dissertation study and, for this reason, are explained in more detail in the methods chapter 4. This preliminary study helped to motivate several of the changes made to the dissertation study research questions and methodology.

A.1 Study Protocol

In this study, we used Amazon Mechanical Turk (AMT) to recruit participants and conduct the entirety of the study. The study protocol was a follows. First, participants were directed to a workflow interface page, shown in Figure A.1. In this interface, participants could view an outline of the study steps, but were only able to click on the current active step highlighted in blue. The first step “Demographics Questionnaire” included a consent form and demographics questionnaire. The next step “Task Description” presented them with the following learning-oriented search task they would be asked to complete during the search session:

> *Scenario*: Every Tuesday night you and your friend cook dinner together. This particular Tuesday night you are making a big salad. When you go to the fridge to get the lettuce, you notice that the leaves have become wilted. Your friend tells you she remembers reading about a cool trick that rehydrates lettuce by soaking it in water. You both decide to try it out. After soaking the lettuce in the water for about an hour, you do indeed notice that the lettuce looks sturdy and revived. Later that night, you decide to learn more about this process. In your research, you come across the principles of diffusion and osmosis.
>
> *Task*: Gather information and **learn everything you can about the concepts of *diffusion* and *osmosis***. Determine which concept (*diffusion* or *osmosis*) is most useful in explaining how lettuce can be rehydrated when soaked in water. Be prepared to explain your answer by providing a well-reasoned, logical argument.

Next, participants were asked to complete a pre-task questionnaire that asked questions about their prior knowledge of the task topic, task difficulty, and a priori determinability (specific questions are shown in the methods chapter in Table 4.2).

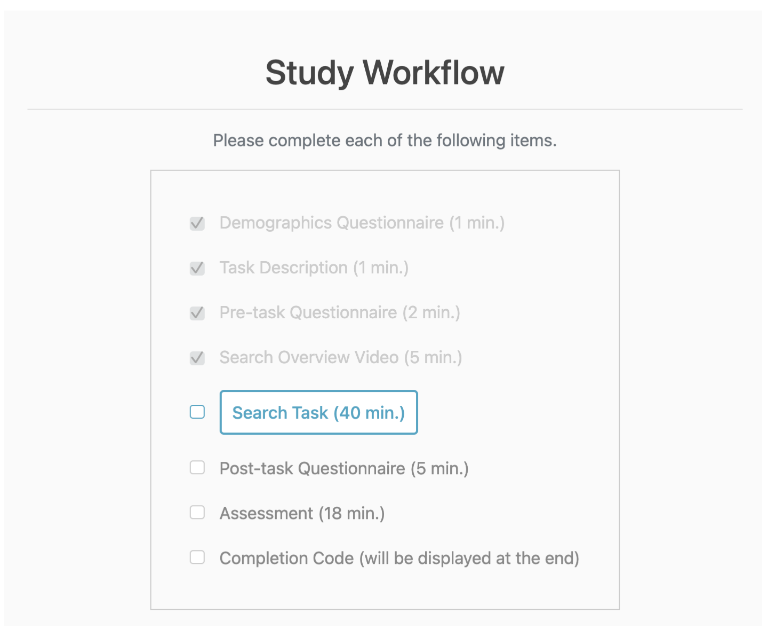

Figure A.1 Workflow interface for Amazon Mechanical Turk workers to navigate through the study (current active step “Search Task” is highlighted in blue, all other steps are not clickable).

After completing the pre-task questionnaire, the next step presented participants with an instructional video that differed based on experimental condition. Participants in the SelfSetSubgoals condition watched a video that explained the functionality of the Subgoal Manager with blank subgoals and how to develop subgoals with ideal qualities (i.e., specific time, action, standard and criteria from McCardle et al. [39]). Participants in the AssignedSubgoals condition watched a video that explained the functionality of the Subgoal Manager with pre-populated subgoals and explained that each subgoal incorporated ideal qualities and described the ideal qualities. Participants in the NoSubgoals condition watched a video that explained the functionality of the text editor they could use to take notes and save information as they searched. The NoSubgoals condition did *not* inform participants about subgoals. Full details on instructional videos are provided in Section 4.4.

After the participants watched the video, they clicked the "Search Task" button. This directed them to an intermediary page to remind them of their objectives before beginning the search session—(1) before searching, open the search tool specific to the experimental condition (i.e., Subgoal Manager or text editor); (2) use the search tool to take notes and save information while searching; and (3) limit of 40 minutes to search. Then, participants were sent to the search interface (shown in the methods chapter in Figure 4.3). During the search session, participants browsed the open web and used the search tool to take notes and save information as they searched. When they were done, participants clicked the "Done with search task" button, provided in the search interface, to proceed to the next step.

After completing the search session, participants completed a post-task questionnaire with items about task difficulty, support of SRL processes, and engagement. Finally, participants completed the ODCA, an 18 item multiple-choice learning assessment from Fisher [1] (details on the ODCA are provided in Chapter 4). Questions on the ODCA are two-tiered, meaning that questions can be considered as pairs of questions. The initial question in the pair asks about some declarative bit of information regarding diffusion and/or osmosis, while the follow-up question asks about a participant's reasoning for selecting the initial answer. Both questions in the pair provide multiple-choice answer options. Participants were given up to 18 minutes to complete the assessment (a maximum of 2 minutes per pair of questions). The study took about 1 hour to complete. Participants were paid $19.00 USD for their participation.

A.2 Results

One-way ANOVA analyses were conducted to compare the effect of the three experimental conditions on post-task variables. First, I discuss the results of ANOVA analysis on post-task ODCA learning assessment scores. Then, I discuss the statistically significant differences in means from ANOVA analyses of post-task questionnaire variables.

A.2.1 Learning Assessment Scores

A one-way ANOVA was performed to compare the effect of three different experimental conditions (i.e., SelfSetSubgoals, AssignedSubgoals, NoSubgoals) on ODCA learning assessment scores. There was no statistically significant difference in mean exam scores on the ODCA between groups. Although not statistically significant, mean scores on the exam were lowest in the AssignedSubgoals condition (M = 10.12, SD = 2.93) and highest in the SelfSetSubgoals condition (M = 11.35, SD = 3.15).

The ODCA assessment contains 16 questions that are organized as 8 pairs. Each pair is associated with a *declarative* question that tests knowledge of a specific concept and a *reasoning* question that tests the participant's justification for their answer to the declarative question (further details of the ODCA learning assessment are discussed in Chapter 4). A one-way ANOVA showed a statistically significant difference in scores on a particular question pair (where participants got *both* the declarative and reasoning questions correct) between at least two groups. Questions 7 and 8 from the ODCA showed a statistically significant difference in scores between at least two groups ($F(2, 118) = [3.582]$, $p = 0.03$). Tukey's HSD Test for multiple comparisons found questions 7 and 8 scores were significantly different between AssignedSubgoals and SelfSetSubgoals (p = 0.05). Participants in SelfSetSubgoals scored on average 0.15 points higher than those in the AssignedSubgoals condition. Additionally, questions 9 and 10 from the ODCA showed a borderline difference in scores between at least two groups ($F(2, 118) = [2.957]$, $p = 0.0558$). Tukey's HSD Test for multiple comparisons found questions 9 and 10 scores were significantly different between AssignedSubgoals and SelfSetSubgoals ($p = 0.05$). Participants in SelfSetSubgoals scored on average 0.26 points higher than those in the AssignedSubgoals condition.

Table A.1 Post-task questionnaire variables with significant differences between conditions.

Variable	$\boldsymbol{F(2{,}119)}$	p**-value**	**Condition**	**Mean, SD**
			SelfSetSubgoals	2.60, 1.32
Difficulty	13.359	$p < 0.001$	AssignedSubgoals	3.19, 1.49
			NoSubgoals	2.04, 0.93
			SelfSetSubgoals	6.14, 0.74
Planning (SRL)	64.92	$p < 0.001$	AssignedSubgoals	6.26, 0.92
			NoSubgoals	3.68, 1.60
			SelfSetSubgoals	6.02, 0.95
Progress (SRL)	8.196	$p < 0.001$	AssignedSubgoals	6.28, 0.92
			NoSubgoals	5.18, 1.62
			SelfSetSubgoals	4.84, 1.33
Adapting (SRL)	4.145	$p = 0.018$	AssignedSubgoals	4.73, 1.66
			NoSubgoals	3.94, 1.63
			SelfSetSubgoals	4.81, 1.46
Aesthetic Appeal	7.075	$p = 0.001$	AssignedSubgoals	5.12, 1.49
			NoSubgoals	3.98, 1.35
			SelfSetSubgoals	6.00, 1.23
Perceived Usability	8.196	$p = 0.006$	AssignedSubgoals	5.41, 1.54
			NoSubgoals	6.32, 0.84
			SelfSetSubgoals	5.18, 1.29
Focused Attention	7.761	$p < 0.001$	AssignedSubgoals	5.46, 1.33
			NoSubgoals	4.27, 1.75

A.2.2 Participant Perceptions

A one-way ANOVA was performed to compare the effect of three different experimental conditions (i.e., SelfSetSubgoals, AssignedSubgoals, NoSubgoals) on post-task questionnaire perceptions of difficulty, SRL support, and engagement. Table A.1 summarizes these results, with means and standard deviations by experimental condition.

Difficulty: A one-way ANOVA was performed to compare the effect of three different experimental conditions (i.e., SelfSetSubgoals, AssignedSubgoals, NoSubgoals) on perceptions of difficulty.

Perceptions of difficulty showed a statistically significant difference between at least two groups ($F(2,\,119) = [13.359]$, $p < 0.001$). Tukey's HSD Test for multiple comparisons found perceptions of difficulty were significantly different between NoSubgoals and AssignedSubgoals ($p < 0.001$). Participants in NoSubgoals rated difficulty on average 1.14 points less than those in the AssignedSubgoals condition.

SRL: A one-way ANOVA was performed to compare the effect of three different experimental conditions (i.e., SelfSetSubgoals, AssignedSubgoals, NoSubgoals) on perceptions of support for SRL.

Perceptions of support for planning showed a statistically significant difference between at least two groups ($F(2,\ 119) = [64.92]$, $p < 0.001$). Tukey's HSD Test for multiple comparisons found perceptions of support for planning were significantly different between NoSubgoals and SelfSetSubgoals ($p < 0.001$) and NoSubgoals and AssignedSubgoals ($p < 0.001$). Participants in SelfSetSubgoals rated support for planning on average 2.43 points higher than those in the NoSubgoals condition and participants in NoSubgoals rated support for planning on average 2.57 points less than those in the AssignedSubgoals condition.

Perceptions of support for evaluating progress and organizing showed a statistically significant difference between at least two groups ($F(2,\ 119) = [8.196]$, $p < 0.001$). Tukey's HSD Test for multiple comparisons found perceptions of support for evaluating progress and organizing were significantly different between NoSubgoals and SelfSetSubgoals ($p = 0.014$) and NoSubgoals and AssignedSubgoals ($p < 0.001$). Participants in SelfSetSubgoals rated support for evaluating progress and organizing on average 0.62 points higher than those in the NoSubgoals condition and participants in NoSubgoals rated support for evaluating progress and organizing on average 0.83 points less than those in the AssignedSubgoals condition.

Perceptions of support for adapting their approach showed a statistically significant difference between at least two groups ($F(2,\ 119) = [4.145]$, $p = 0.018$). Tukey's HSD Test for multiple comparisons found perceptions of support for adapting their approach were significantly different between NoSubgoals and SelfSetSubgoals ($p = 0.027$). Participants in SelfSetSubgoals rated support for adapting on average 0.89 points higher than those in the NoSubgoals condition.

Engagement: A one-way ANOVA was performed to compare the effect of three different experimental conditions (i.e., SelfSetSubgoals, AssignedSubgoals, NoSubgoals) on perceptions of engagement.

Perceptions of aesthetic appeal showed a statistically significant difference between at least two groups ($F(2,\ 119) = [7.075]$, $p < 0.001$). Tukey's HSD Test for multiple comparisons found perceptions of aesthetic appeal were significantly different between NoSubgoals and SelfSet-Subgoals ($p = 0.03$) and NoSubgoals and AssignedSubgoals ($p = 0.001$). Participants in SelfSetSubgoals rated on average 0.81 points higher than those in the NoSubgoals condition and participants in NoSubgoals rated on average 1.16 points less than those in the AssignedSubgoals condition.

Perceptions of perceived usability showed a statistically significant difference between at least two groups ($F(2, 119) = [8.196]$, p $= 0.006$). Tukey's HSD Test for multiple comparisons found perceptions of perceived usability were significantly different NoSubgoals and AssignedSubgoals ($p < 0.004$). Participants in NoSubgoals rated on average 0.89 points higher than those in the AssignedSubgoals condition.

Perceptions of focused attention showed a statistically significant difference between at least two groups ($F(2, 119) = [7.761]$, $p < 0.001$). Tukey's HSD Test for multiple comparisons found perceptions of focused attention were significantly different between NoSubgoals and SelfSetSubgoals ($p = 0.01$) and NoSubgoals and AssignedSubgoals ($p = 0.001$). Participants in SelfSetSubgoals rated on average 0.92 points higher than those in the NoSubgoals condition and participants in NoSubgoals rated on average 1.22 points less than those in the AssignedSubgoals condition.

A.3 Discussion

This preliminary study was helpful for several reasons. First, the preliminary study helped me to better understand the internal consistency or reliability of novel questionnaire items. Cronbach's alpha and exploratory factor analyses of questions allowed for combining of certain questions for analysis. For example, prior knowledge questions all loaded as a single factor that could be used as a covariate analyzing significant differences of learning scores between experimental conditions. Additionally, Cronbach's alpha and exploratory factor analyses prompted editing and refinement of questions. For example, half of the SRL questions designed to measure "organizing information" loaded with "evaluating progress" questions, while the other half loaded with "connecting ideas" questions. For this reason, organizing will no longer be represented as a separate section. Some of the questions have been eliminated and some have been absorbed into evaluating and connecting sections.

Second, the preliminary study demonstrated that the ODCA assessment alone may be too specific in isolation to capture all learning that occurred during the search session. While results showed that scores on particular question pairs were significantly different across conditions, it would have been useful to have an additional open-ended assessment to capture a broader picture of *all* that participants learned across experimental conditions. For this reason, an open-ended question has been added to my dissertation study (shown in Figure 4.5).

Finally, the preliminary study demonstrated significant differences between experimental conditions in perceptions of SRL process support in the areas of planning, evaluating progress, and adapting. Participants in the SelfSetSubgoals and AssignedSubgoals conditions found that the Subgoal Manager supported them in planning and evaluating subgoal progress more than participants in the NoSubgoals condition. Also, participants in the SelfSetSubgoals condition found that the Subgoal Manager supported them in adapting their approach. These results indicate that participants did perceive higher levels of support of macro-level SRL processes in conditions with the Subgoal Manager. However, perceptions of SRL support are quite limited and clearly much different than capturing and describing *actual* SRL processes from observation. Greene et al. [153] states, "Compared to retrospective self-report methods, the real-time collection of these various kinds of trace data, either via direct observation of student behavior, or think-aloud protocol or log-file data, provides greater confidence in the measurement of SRL knowledge, skills, and dispositions, and allows for more sophisticated analysis of their temporal and adaptive nature." [154, p. 2] For this reason, a think-aloud protocol will be used in my dissertation study to investigate the frequency of SRL processes across different conditions. Importantly, such detailed analysis will also afford the ability to understand what types of SRL processes occur and when.

APPENDIX B

Statements From Sentence Coding Guide

Rule 1: Compound Subjects are Multiple Statements

- An instance of compound subjects related to an object or objects are considered multiple statements.

1. Subjects x and y are object z.
    - Statement 1: “Subject y is object z.”
    - Statement 2: “Subject x is object z.”

Example:

- 1. “Diffusion and osmosis are forms of passive transport.”
        - Statement 1: “Diffusion is a form of passive transport”
        - Statement 2: “Osmosis is a form of passive transport”

    “Both diffusion and osmosis involve particles moving from areas of high concentration to areas of lower concentration.”

    2. - Statement 1: “Diffusion involves particles moving from areas of high concentration to areas of lower concentration”
        - Statement 2: “Osmosis involves particles moving from areas of high concentration to areas of lower concentration”

Rule 2: Compound Objects are Multiple Statements

- An instance of subjects related to compound objects are considered multiple statements.

1. Subjects x and y are object z and object alpha.
    1. Statement 1: “Subject x is object z.”
    2. Statement 2: “Subject y is object z.”
    3. Statement 3: “Subject x is object alpha.”
    4. Statement 4: “Subject y is object alpha.”

Examples:

- 1. "There are two types of diffusion; simple and facilitated diffusion."

    i. <u>Statement 1</u>: "One type of diffusion is simple diffusion"

    ii. <u>Statement 2</u>: "One type of diffusion is facilitated diffusion"

  "In the body, diffusion is important for the transport of nutrients and energy."

  2. i. <u>Statement 1</u>: "Diffusion is important for the transport of nutrients"

    ii. <u>Statement 2</u>: "Diffusion is important for the transport of energy"

Rule 3: Exemplifying Statements are Statements

- An instance of an example of a statement is considered an additional statement.

  1. Subject x is a y exemplified by z.

    i. <u>Statement 1</u>: "Subject x is a y"

    ii. <u>Statement 2</u>: "Object y is exemplified by z"

  Examples:

- 1. "Diffusion can be affected by multiple factors such as temperature."

    i. <u>Statement 1</u>: "Diffusion can be affected by multiple factors"

    ii. <u>Statement 2</u>: "One factor diffusion is affected by is temperature"

Rule 4: Definitions are Multiple Statements

- An instance of a definition is considered a single statement (rather than separated into multiple statements).

  1. Object w is defined as x from y to z.

    i. <u>Statement 1</u>: "Object w involves x"

    ii. <u>Statement 2</u>: "Object w involves y"

    iii. <u>Statement 3</u>: "Object w involves z"

Example:

- 1. "Simple diffusion involves small molecules passively moving through a membrane from areas of high to low concentration."

    i. Statement 1: "Simple diffusion involves small molecules"

    ii. Statement 2: "Simple diffusion involves passive movement."

    iii. Statement 3: "Simple diffusion involves a membrane."

    iv. Statement 4: "Simple diffusion involves movement from areas of high to low concentration."

  "Osmosis...focuses on the movement of water molecules (or other solvent but typically water) from dilute to concentrated regions across a semipermeable membrane."

  2. i. Statement 1: "Osmosis involves the movement of a solvent"

    ii. Statement 2: "Osmosis involves water molecules or other solvent"

    iii. Statement 3: "Osmosis involves movement from dilute to concentrated regions"

    iv. Statement 4: "Osmosis involves movement across a semipermeable membrane"

Rule 5: Definitions within Statements are Multiple Statements

- An instance of a definition stated within a statement is considered an additional statement.

  1. Subject x is an object y (defined as z).

    i. Statement 1: "Subject x is an object y"

    ii. Statement 2: "Object y is defined as z"

Example:

- 1. "Overall diffusion includes simple diffusion and facilitated diffusion (movement of particles from areas of high to low concentration with and without the help of carrier molecules, respectively)."

    i. Statement 1: "Diffusion includes simple diffusion" *(by Rule 2)*

    ii. Statement 2: "Diffusion includes facilitated diffusion" *(by Rule 2)*

    iii. Statement 3: "Simple diffusion involves the movement of particles"

iv. Statement 4: "Simple diffusion involves movement from areas of high to low concentration"

v. Statement 5: "Simple diffusion involves the help of carrier molecules"

vi. Statement 6: "Facilitated diffusion involves the movement of particles"

vii. Statement 7: "Facilitated diffusion involves movement from areas of high to low concentration"

viii. Statement 8: "Facilitated diffusion involves movement without the help of carrier molecules"

Rule 6: Statement Implied by Statement is not a Statement

- An instance of a statement implied within a statement is not considered an additional statement.
- Examples:
  1. "There are three different types of solutions used to describe osmosis; isotonic (concentration is equal in and outside of the cell), hypertonic (concentration is higher inside) and hypotonic."
     i. Stating that there are three different types of solutions is not an additional statement because it is *implied* as there are three types listed (i.e., isotonic, hypertonic, and hypotonic); therefore this example statement has five statements (rather than six statements):
        1. Statement 1: "Isotonic solutions are used to describe osmosis." *(by Rule 2)*
        2. Statement 2: "Hypertonic solutions are used to describe osmosis." *(by Rule 2)*
        3. Statement 3: "Hypotonic solutions are used to describe osmosis." *(by Rule 2)*
        4. Statement 4: "isotonic means the concentration is equal in and outside of the cell" *(by Rule 3)*
        5. Statement 5: "hypertonic means the concentration is higher inside the cell" *(by Rule 3)*
        6. Statement 6: "There are three types of solutions to describe osmosis"

APPENDIX C

Statement List Coding Guide

Please go through each statement and assign one statement from the list of unique statements. When going through this process:

1. You can use ctrl+f to search through the unique statement list with keywords.
2. Just because a statement has been assigned to a particular participant doesn't mean it won't be assigned again to that participant
   a. Each sentence is considered in isolation
      i. For example:
         1. Sentence 1 "Diffusion is the movement of particles from areas of high concentration to low concentration."
            - Sentence 1 is broken apart into:
              - Statement 1 **diff_hightolow** *(Diffusion involves movement from a region of high concentration to a region of low concentration.)*
              - Statement 2, etc...

            Sentence 2 "An example of diffusion is a perfume smell moving from an area of high concentration to an area of low concentration throughout a room."
         2. - Sentence 2 is broken apart into:
              - Statement 1 **diff_hightolow** *(Diffusion involves movement from a region of high concentration to a region of low concentration.)*
              - Statement 2, etc...

   Select the most specific statement
3. a. For example:
      i. Sentence 3 "Osmosis requires a semipermeable membrane."
         1. Sentence 3 is assigned:
            - **osm_semiperm_req** (*Osmosis requires a semipermeable membrane.*), rather than **osm_semiperm** (*Osmosis involves a semipermeable membrane.*)

REFERENCES

[1] K. M. Fisher, K. S. Williams, and J. E. Lineback, "Osmosis and Diffusion Conceptual Assessment," *CBE—Life Sciences Education*, vol. 10, pp. 418–429, Dec. 2011. Publisher: American Society for Cell Biology (lse). (document), 3, 3.1, 4.8, 4.4, A.1

[2] P. H. Winne and A. F. Hadwin, "Studying as self-regulated engagement in learning," in *Metacognition in educational theory and practice*, 1998. (document), 1, 2.2, 2.2.1, 3, 4.7.3, 4.8, 6.4

[3] H. O'Brien, P. Cairns, and M. Hall, "A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form | Elsevier Enhanced Reader," 2018. (document), 4.7.3, 4.10

[4] S. Y. Rieh, K. Collins-Thompson, P. Hansen, and H.-J. Lee, "Towards searching as a learning process: A review of current perspectives and future directions," *Journal of Information Science*, vol. 42, pp. 19–34, Feb. 2016. 1

[5] C. Eickhoff, J. Teevan, R. White, and S. Dumais, "Lessons from the journey: a query log analysis of within-session learning," in *Proceedings of the 7th ACM international conference on Web search and data mining - WSDM '14*, (New York, New York, USA), pp. 223–232, ACM Press, 2014. 1

[6] P. Bailey, L. Chen, and S. Grosenick, "User task understanding: a web search engine perspective," *NII Shonan workshop*, p. 17, Oct. 2012. 1

[7] K. Collins-Thompson, P. Hansen, and C. Hauff, "Search as Learning (Dagstuhl Seminar 17092)," 2017. 1, 2.1

[8] J. Allan, B. Croft, A. Moffat, and M. Sanderson, "Frontiers, challenges, and opportunities for information retrieval: Report from SWIRL 2012 the second strategic workshop on information retrieval in Lorne," *ACM SIGIR Forum*, vol. 46, p. 2, May 2012. 1, 2.1

[9] L. Freund, J. He, J. Gwizdka, N. Kando, P. Hansen, and S. Y. Rieh, "Searching As Learning (SAL) Workshop 2014," in *Proceedings of the 5th Information Interaction in Context Symposium*, IIiX '14, (New York, NY, USA), pp. 7–7, ACM, 2014. 1

[10] J. Gwizdka, P. Hansen, C. Hauff, J. He, and N. Kando, "Search As Learning (SAL) Workshop 2016," in *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '16, (New York, NY, USA), pp. 1249–1250, ACM, 2016. 1

[11] T. Willoughby, S. A. Anderson, E. Wood, J. Mueller, and C. Ross, "Fast searching for information on the Internet to use in a learning context: The impact of domain knowledge," *Computers & Education*, vol. 52, pp. 640–648, Apr. 2009. 1, 2.1, 2.1.1, 2.5

[12] H. L. O'Brien, A. Kampen, A. W. Cole, and K. Brennan, "The Role of Domain Knowledge in Search as Learning," in *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, CHIIR '20, (Vancouver BC, Canada), pp. 313–317, Association for Computing Machinery, Mar. 2020. 1, 2.1, 2.1.1, 2.5

[13] N. Roy, F. Moraes, and C. Hauff, "Exploring Users' Learning Gains within Search Sessions," in *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, CHIIR '20, (Vancouver BC, Canada), pp. 432–436, Association for Computing Machinery, Mar. 2020. 1, 2.1, 2.1.1, 2.5

[14] G. Pardi, J. von Hoyer, P. Holtz, and Y. Kammerer, "The Role of Cognitive Abilities and Time Spent on Texts and Videos in a Multimodal Searching as Learning Task," in *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, (Vancouver BC Canada), pp. 378–382, ACM, Mar. 2020. 1, 2.1.1, 2.5

[15] S. Ghosh, M. Rath, and C. Shah, "Searching As Learning: Exploring Search Behavior and Learning Outcomes in Learning-related Tasks," in *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, CHIIR '18, (New York, NY, USA), pp. 22–31, ACM, 2018. 1, 2.1.2, 2.5

[16] R. Kalyani and U. Gadiraju, "Understanding User Search Behavior Across Varying Cognitive Levels," in *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, HT '19, (New York, NY, USA), pp. 123–132, Association for Computing Machinery, Sept. 2019. 1, 2.1.2, 2.5

[17] H. Liu, C. Liu, and N. J. Belkin, "Investigation of users' knowledge change process in learning-related search tasks," *Proceedings of the Association for Information Science and Technology*, vol. 56, no. 1, pp. 166–175, 2019. 1, 2.1.2, 2.5

[18] Y. Kammerer, R. Nairn, P. Pirolli, and E. H. Chi, "Signpost from the masses: learning effects in an exploratory social tag search browser," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, (Boston, MA, USA), pp. 625–634, Association for Computing Machinery, Apr. 2009. 1, 2.1, 2.1.3, 2.5

[19] L. Freund, R. Kopak, and H. O'Brien, "The effects of textual environment on reading comprehension: Implications for searching as learning," *Journal of Information Science*, vol. 42, pp. 79–93, Feb. 2016. Publisher: SAGE Publications Ltd. 1, 2.1, 2.1.3, 2.5

[20] N. Roy, M. V. Torre, U. Gadiraju, D. Maxwell, and C. Hauff, "Note the Highlight: Incorporating Active Reading Tools in a Search as Learning Environment," in *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, CHIIR '21, (New York, NY, USA), pp. 229–238, Association for Computing Machinery, Mar. 2021. 1, 2.1, 2.1.3, 2.5

[21] A. Câmara, N. Roy, D. Maxwell, and C. Hauff, "Searching to Learn with Instructional Scaffolding," in *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, CHIIR '21, (New York, NY, USA), pp. 209–218, Association for Computing Machinery, Mar. 2021. 1, 2.1.3, 2.5

[22] R. Syed and K. Collins-Thompson, "Optimizing search results for human learning goals," *Information Retrieval Journal*, vol. 20, pp. 506–523, Oct. 2017. 1, 2.1.3, 2.1.4, 2.5

[23] N. Weingart and C. Eickhoff, "Retrieval Techniques for Contextual Learning," *SAL @ SIGIR*, p. 5, 2016. 1, 2.1.3, 2.5

[24] U. Gadiraju, R. Yu, S. Dietze, and P. Holtz, "Analyzing Knowledge Gain of Users in Informational Search Sessions on the Web," in *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, CHIIR '18, (New York, NY, USA), pp. 2–11, ACM, 2018. 1, 2.1, 2.1.4, 2.5, 3.3, 4.9, 6.3

[25] K. Collins-Thompson, S. Y. Rieh, C. C. Haynes, and R. Syed, "Assessing Learning Outcomes in Web Search: A Comparison of Tasks and Query Strategies," in *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, CHIIR '16, (New York, NY, USA), pp. 163–172, ACM, 2016. 1, 2.1, 2.1.4, 2.5, 3.3

[26] R. Yu, U. Gadiraju, P. Holtz, M. Rokicki, P. Kemkes, and S. Dietze, "Predicting User Knowledge Gain in Informational Search Sessions," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, (New York, NY, USA), pp. 75–84, ACM, 2018. 1, 2.1, 2.1.4, 2.5, 3.3, 4.9, 6.3

[27] C. Liu and X. Song, "How do Information Source Selection Strategies Influence Users' Learning Outcomes'," in *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, CHIIR '18, (New York, NY, USA), pp. 257–260, Association for Computing Machinery, Mar. 2018. 1, 2.1, 2.1.4, 2.5, 3.3

[28] Y. Lu and I.-H. Hsiao, "Personalized Information Seeking Assistant (PiSA): from programming information seeking to learning," *Information Retrieval Journal*, vol. 20, pp. 433–455, Oct. 2017. 1, 2.1, 2.1.4, 3.3

[29] M. Abualsaud, "Learning Factors and Determining Document-level Satisfaction In Search-as-Learning," Master's thesis, University of Waterloo, Waterloo, Ontario, Canada, 2017. 1, 2.1, 2.1.4, 2.5, 3.3

[30] N. Bhattacharya and J. Gwizdka, "Measuring Learning During Search: Differences in Interactions, Eye-Gaze, and Semantic Similarity to Expert Knowledge," in *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, (Glasgow Scotland UK), pp. 63–71, ACM, Mar. 2019. 1, 2.1, 2.1.4, 2.5, 3.3

[31] S. Palani, Z. Ding, S. MacNeil, and S. P. Dow, "The "Active Search" Hypothesis: How Search Strategies Relate to Creative Learning," in *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, CHIIR '21, (New York, NY, USA), pp. 325–329, Association for Computing Machinery, Mar. 2021. 1, 2.1, 2.1.4, 2.5, 3.3

[32] M. Boekaerts, M. Zeidner, P. R. Pintrich, and P. R. Pintrich, *Handbook of Self-Regulation.* San Diego, UNITED STATES: Elsevier Science & Technology, 1999. 1, 2.2, 2.2.1, 3

[33] T. Sitzmann and K. Ely, "A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go," *Psychological Bulletin*, vol. 137, no. 3, pp. 421–442, 2011. Place: US Publisher: American Psychological Association. 1, 2.2, 2.2.1, 2.3.1, 3, 3.1

[34] D. H. Schunk and C. W. Swartz, "Goals and Progress Feedback: Effects on Self-Efficacy and Writing Achievement," *Contemporary Educational Psychology*, vol. 18, pp. 337–354, July 1993. 1, 2.2, 3

[35] B. J. Zimmerman and D. H. Schunk, *Handbook of self-regulation of learning and performance.* Handbook of self-regulation of learning and performance, New York, NY, US: Routledge/Taylor & Francis Group, 2011. Pages: xiv, 484. 1, 2.2, 3

[36] J. A. Greene and R. Azevedo, "A Theoretical Review of Winne and Hadwin's Model of Self-Regulated Learning: New Perspectives and Directions," *Review of Educational Research*, vol. 77, pp. 334–372, Sept. 2007. Publisher: American Educational Research Association. 1, 2.2.1

[37] J. A. Greene, C. M. Bolick, W. P. Jackson, A. M. Caprino, C. Oswald, and M. McVea, "Domain-specificity of self-regulated learning processing in science and history," *Contemporary Educational Psychology*, vol. 42, pp. 111–128, July 2015. 1, 2.2.2, 4.12, 7.1

[38] P. H. Winne and N. E. Perry, "Measuring self-regulated learning," in *Handbook of self-regulation*, pp. 531–566, San Diego, CA, US: Academic Press, 2000. 1, 2.2.1, 3

[39] L. McCardle, E. A. Webster, A. Haffey, and A. F. Hadwin, "Examining students' self-set goals for self-regulated learning: Goal properties and patterns," *Studies in Higher Education*, vol. 42, pp. 2153–2169, Nov. 2017. Publisher: Routledge _eprint: https://doi.org/10.1080/03075079.2015.1135117. 1, 2.3, 2.3.2, 3.4, 4.4, 6.4, A.1

[40] B. J. Jansen, D. Booth, and B. Smith, "Using the taxonomy of cognitive learning to model online searching," *Information Processing & Management*, vol. 45, pp. 643–663, Nov. 2009. 2.1

[41] W.-C. Wu, D. Kelly, A. Edwards, and J. Arguello, "Grannies, Tanning Beds, Tattoos and NASCAR: Evaluation of Search Tasks with Varying Levels of Cognitive Complexity," in *Proceedings of the 4th Information Interaction in Context Symposium*, IIIX '12, (New York, NY, USA), pp. 254–257, ACM, 2012. event-place: Nijmegen, The Netherlands. 2.1

[42] D. Kelly, J. Arguello, A. Edwards, and W.-c. Wu, "Development and Evaluation of Search Tasks for IIR Experiments Using a Cognitive Complexity Framework," in *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, ICTIR '15, (New York, NY, USA), pp. 101–110, ACM, 2015. 2.1

[43] R. Capra, J. Arguello, A. Crescenzi, and E. Vardell, "Differences in the Use of Search Assistance for Tasks of Varying Complexity," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, (New York, NY, USA), pp. 23–32, ACM, 2015. 2.1

[44] D. Demaree, H. Jarodzka, S. Brand-Gruwel, and Y. Kammerer, "The Influence of Device Type on Querying Behavior and Learning Outcomes in a Searching as Learning Task with a Laptop or Smartphone," in *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, CHIIR '20, (New York, NY, USA), pp. 373–377, Association for Computing Machinery, Mar. 2020. 2.1, 2.5, 3.3

[45] L. W. Anderson, D. R. Krathwohl, P. W. Airasian, K. A. Cruikshank, R. E. Mayer, P. R. Pintrich, J. Raths, and M. C. Wittrock, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged Edition.* New York: Pearson, 1 edition ed., Dec. 2000. 2.1.2

[46] J. Liu, N. J. Belkin, X. Zhang, and X. Yuan, "Examining users' knowledge change in the task completion process," *Information Processing & Management*, vol. 49, pp. 1058–1074, Sept. 2013. 2.1.2, 7.3

[47] B. J. Zimmerman and M. M. Pons, "Development of a Structured Interview for Assessing Student Use of Self-Regulated Learning Strategies," *American Educational Research Journal*, vol. 23, pp. 614–628, Jan. 1986. Publisher: American Educational Research Association. 2.2, 7.1

[48] B. J. Zimmerman and M. Martinez-Pons, "Construct validation of a strategy model of student self-regulated learning," *Journal of Educational Psychology*, vol. 80, no. 3, pp. 284–290, 1988. Place: US Publisher: American Psychological Association. 2.2, 7.1

[49] D. H. Schunk, "Sequential attributional feedback and children's achievement behaviors," *Journal of Educational Psychology*, vol. 76, no. 6, pp. 1159–1169, 1984. Place: US Publisher: American Psychological Association. 2.2, 7.1

[50] D. H. Schunk, "Modeling and attributional effects on children's achievement: A self-efficacy analysis," *Journal of Educational Psychology*, vol. 73, no. 1, pp. 93–105, 1981. Place: US Publisher: American Psychological Association. 2.2, 7.1

[51] B. J. Zimmerman, "Self-Regulated Learning and Academic Achievement: An Overview," *Educational Psychologist*, vol. 25, pp. 3–17, Jan. 1990. 2.2

[52] B. J. Zimmerman, "Special issue on self-regulated learning," *Contemporary Educational Psychology*, vol. 11, no. 4, pp. 305–427, 1986. 2.2

[53] D. C. Moos and R. Azevedo, "Monitoring, planning, and self-efficacy during learning with hypermedia: The impact of conceptual scaffolds," *Computers in Human Behavior*, vol. 24, pp. 1686–1706, July 2008. 2.2

[54] P. H. Winne, "Self-regulated learning viewed from models of information processing," in *Self-regulated learning and academic achievement: Theoretical perspectives, 2nd ed*, pp. 153–189, Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2001. 2.2

[55] D. H. Schunk, "Self-Regulation Through Goal Setting," *ERIC Digest*, p. 2, 2001. 2.2, 2.3

[56] P. R. Pintrich, "Chapter 14 - The Role of Goal Orientation in Self-Regulated Learning," in *Handbook of Self-Regulation* (M. Boekaerts, P. R. Pintrich, and M. Zeidner, eds.), pp. 451–502, San Diego: Academic Press, Jan. 2000. 2.2.1

[57] B. J. Zimmerman, "Chapter 2 - Attaining Self-Regulation: A Social Cognitive Perspective," in *Handbook of Self-Regulation* (M. Boekaerts, P. R. Pintrich, and M. Zeidner, eds.), pp. 13–39, San Diego: Academic Press, Jan. 2000. 2.2.1

[58] B. J. Zimmerman, "Becoming a Self-Regulated Learner: An Overview," *Theory Into Practice*, vol. 41, pp. 64–70, May 2002. 2.2.1

[59] M. Puustinen and L. Pulkkinen, "Models of Self-regulated Learning: A review," *Scandinavian Journal of Educational Research*, vol. 45, pp. 269–286, Sept. 2001. Publisher: Routledge _eprint: https://doi.org/10.1080/00313830120074206. 2.2.1

[60] A. Bandura, *Social foundations of thought and action: A social cognitive theory.* Social foundations of thought and action: A social cognitive theory, Englewood Cliffs, NJ, US: Prentice-Hall, Inc, 1986. Pages: xiii, 617. 2.2.1

[61] J. Kuhl, "From Cognition to Behavior: Perspectives for Future Research on Action Control," in *Action Control: From Cognition to Behavior* (J. Kuhl and J. Beckmann, eds.), SSSP Springer Series in Social Psychology, pp. 267–275, Berlin, Heidelberg: Springer, 1985. 2.2.1

[62] G. van den Boom, F. Paas, and J. J. G. van Merriënboer, "Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes," *Learning and Instruction*, vol. 17, pp. 532–548, Oct. 2007. 2.2.1

[63] I. Glogger, R. Schwonke, L. Holzäpfel, M. Nückles, and A. Renkl, “Learning strategies assessed by journal writing: Prediction of learning outcomes by quantity, quality, and combinations of learning strategies,” *Journal of Educational Psychology*, vol. 104, no. 2, pp. 452–468, 2012. Place: US Publisher: American Psychological Association. 2.2.1

[64] R. Santhanam, S. Sasidharan, and J. Webster, “Using Self-Regulatory Learning to Enhance E-Learning-Based Information Technology Training,” *Information Systems Research*, vol. 19, pp. 26–47, Mar. 2008. Publisher: INFORMS. 2.2.1

[65] M. Bannert, M. Hildebrand, and C. Mengelkamp, “Effects of a metacognitive support device in learning environments,” *Computers in Human Behavior*, vol. 25, pp. 829–835, July 2009. 2.2.1

[66] S. Kistner, K. Rakoczy, B. Otto, C. Dignath-van Ewijk, G. Büttner, and E. Klieme, “Promotion of self-regulated learning in classrooms: investigating frequency, quality, and consequences for student performance,” *Metacognition and Learning*, vol. 5, pp. 157–171, Aug. 2010. 2.2.1

[67] R. Azevedo, S. Ragan, J. G. Cromley, and S. Pritchett, *Do Different Goal-Setting Conditions Facilitate Students’ Ability to Regulate Their Learning of Complex Science Topics with RiverWeb?* Apr. 2002. 2.2.1, 2.3, 3, 3.2

[68] M. Sobocinski, S. Järvelä, J. Malmberg, M. Dindar, A. Isosalo, and K. Noponen, “How does monitoring set the stage for adaptive regulation or maladaptive behavior in collaborative learning?,” *Metacognition and Learning*, vol. 15, pp. 99–127, Aug. 2020. 2.2.1

[69] P. H. Winne, “Students’ calibration of knowledge and learning processes: Implications for designing powerful software learning environments,” *International Journal of Educational Research*, vol. 41, pp. 466–488, Jan. 2004. 2.2.1

[70] R. Azevedo, J. G. Cromley, and D. Seibert, “Does adaptive scaffolding facilitate students’ ability to regulate their learning with hypermedia?,” *Contemporary Educational Psychology*, vol. 29, pp. 344–370, July 2004. 2.2.1

[71] D. H. Schunk and J. A. Greene, *Handbook of Self-Regulation of Learning and Performance*. Routledge, Sept. 2017. 2.2.1, 3

[72] G. Schraw and R. S. Dennison, “Assessing Metacognitive Awareness,” *Contemporary Educational Psychology*, vol. 19, pp. 460–475, Oct. 1994. 2.2.2

[73] P. R. Pintrich and E. V. de Groot, “Motivational and self-regulated learning components of classroom academic performance,” *Journal of Educational Psychology*, vol. 82, no. 1, pp. 33–40, 1990. Place: US Publisher: American Psychological Association. 2.2.2

[74] C. Weinstein, D. Palmer, and A. Schulte, “Learning and Study Strategies Inventory (LASSI),” *Clearwater, FL: H & H Publishing*, 1987. 2.2.2

[75] P. H. Winne, D. Jamieson-Noel, and K. Muis, “Methodological issues and advances in researching tactics, strategies, and self-regulated learning.,” in *New directions in measures and methods* (P. R. Pintrich and M. L. Maehr, eds.), no. 12 in Advances in motivation and achievement, Amsterdam: JAI, An Imprint of Elsevier Science, 1. ed ed., 2002. 2.2.2

[76] J. A. Greene, K. R. Dellinger, B. B. Tüysüzoğlu, and L.-J. Costa, “A Two-Tiered Approach to Analyzing Self-Regulated Learning Data to Inform the Design of Hypermedia Learning Environments,” in *International Handbook of Metacognition and Learning Technologies* (R. Azevedo

and V. Aleven, eds.), Springer International Handbooks of Education, pp. 117–128, New York, NY: Springer, 2013. 2.2.2, 6.5.3

[77] J. A. Greene, V. M. Deekens, D. Z. Copeland, and S. Yu, "Capturing and modeling self-regulated learning using think-aloud protocols," in *Handbook of self-regulation of learning and performance, 2nd ed*, Educational psychology handbook series, pp. 323–337, New York, NY, US: Routledge/Taylor & Francis Group, 2018. 2.2.2

[78] J. A. Greene, L. A. Hutchison, L.-J. Costa, and H. Crompton, "Investigating how college students' task definitions and plans relate to self-regulated learning processing and understanding of a complex science topic," *Contemporary Educational Psychology*, vol. 37, pp. 307–320, Oct. 2012. 2.2.2

[79] J. A. Greene, D. Z. Copeland, V. M. Deekens, and S. B. Yu, "Beyond knowledge: Examining digital literacy's role in the acquisition of understanding in science," *Computers & Education*, vol. 117, pp. 141–159, Feb. 2018. 2.2.2, 4.12, 7.1

[80] R. Azevedo, "Using Hypermedia as a Metacognitive Tool for Enhancing Student Learning? The Role of Self-Regulated Learning," *Educational Psychologist*, vol. 40, pp. 199–209, Dec. 2005. 2.2.2

[81] E. A. Locke and G. P. Latham, "Building a practically useful theory of goal setting and task motivation: A 35-year odyssey," *American Psychologist*, vol. 57, pp. 705–717, Sept. 2002. Publisher: American Psychological Association. 2.3, 4, 3, 3.4

[82] E. A. Locke and G. P. Latham, *A theory of goal setting & task performance*. A theory of goal setting & task performance, Englewood Cliffs, NJ, US: Prentice-Hall, Inc, 1990. Pages: xviii, 413. 2.3

[83] E. A. Locke and G. P. Latham, *New Developments in Goal Setting and Task Performance*. London, UNITED KINGDOM: Routledge, 2012. 2.3, 2.3.1, 2.3.2

[84] E. A. Locke and G. P. Latham, "New Directions in Goal-Setting Theory," *Current Directions in Psychological Science*, vol. 15, pp. 265–268, Oct. 2006. Publisher: SAGE Publications Inc. 2.3, 3

[85] E. A. Locke and G. P. Latham, "The development of goal setting theory: A half century retrospective," *Motivation Science*, vol. 5, pp. 93–105, June 2019. Publisher: Educational Publishing Foundation. 2.3

[86] E. A. Locke, D.-O. Chah, S. Harrison, and N. Lustgarten, "Separating the effects of goal specificity from goal level," *Organizational Behavior and Human Decision Processes*, vol. 43, pp. 270–287, Apr. 1989. 2.3

[87] J. A. LePine, "Adaptation of teams in response to unforeseen change: Effects of goal difficulty and team composition in terms of cognitive ability and goal orientation," *Journal of Applied Psychology*, vol. 90, pp. 1153–1167, Nov. 2005. Publisher: American Psychological Association Inc. 2.3

[88] P. C. Earley, "Influence of information, choice and task complexity upon goal acceptance, performance, and personal goals," *Journal of Applied Psychology*, vol. 70, pp. 481–491, Aug. 1985. Publisher: American Psychological Association. 2.3

[89] J. R. Hollenbeck and H. J. Klein, "Goal commitment and the goal-setting process: Problems, prospects, and proposals for future research," *Journal of Applied Psychology*, vol. 72, no. 2, pp. 212–220, 1987. Place: US Publisher: American Psychological Association. 2.3

[90] G. P. Latham and G. H. Seijts, "The effects of proximal and distal goals on performance on a moderately complex task," *Journal of Organizational Behavior*, vol. 20, no. 4, pp. 421–429, 1999. 2.3

[91] D. Schunk, "Self-Efficacy and Academic Motivation.," *Educational Psychologist*, vol. 26, no. 3 & 4, pp. 207–231, 1991. 2.3

[92] G. P. Latham and E. A. Locke, "New Developments in and Directions for Goal-Setting Research," *European Psychologist*, vol. 12, pp. 290–300, Jan. 2007. 2.3, 2.3

[93] E. Elliott and C. Dweck, "Goals: An approach to motivation and achievement.," *Journal of Personality and Social Psychology*, vol. 54, no. 1, pp. 5–12, 1988. 2.3

[94] G. P. Latham and T. C. Brown, "The Effect of Learning vs. Outcome Goals on Self-Efficacy, Satisfaction and Performance in an MBA Program," *Applied Psychology: An International Review*, vol. 55, no. 4, pp. 606–623, 2006. Place: United Kingdom Publisher: Blackwell Publishing. 2.3, 2.3.1

[95] A. Valle, R. G. Cabanach, J. C. Núnez, J. González-Pienda, S. Rodríguez, and I. Piñeiro, "Multiple goals, motivation and academic learning," *British Journal of Educational Psychology*, vol. 73, no. 1, pp. 71–87, 2003. _eprint: https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1348/000709903762869923. 2.3

[96] G. P. Latham, T. R. Mitchell, and D. L. Dossett, "Importance of participative goal setting and anticipated rewards on goal difficulty and job performance," *Journal of Applied Psychology*, vol. 63, no. 2, pp. 163–171, 1978. Place: US Publisher: American Psychological Association. 2.3, 3

[97] G. P. Latham and L. M. Saari, "Application of social-learning theory to training supervisors through behavioral modeling," *Journal of Applied Psychology*, vol. 64, no. 3, pp. 239–246, 1979. Place: US Publisher: American Psychological Association. 2.3, 3, 3.4

[98] A. Bandura, "Self-Efficacy," in *The Corsini Encyclopedia of Psychology*, pp. 1–3, American Cancer Society, 2010. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470479216.corpsy0836. 2.3

[99] S. S. White and E. A. Locke, "Problems with the pygmalion effect and some proposed solutions," *The Leadership Quarterly*, vol. 11, pp. 389–415, Sept. 2000. 2.3

[100] B. Zimmerman, "Goal Setting: A Key Proactive Source of Academic Self-Regulation," in *Motivation and Self-Regulated Learning: Theory, Research, and Applications*, vol. 267, 2008. 2.3

[101] C.-Y. Chou and N.-B. Zou, "An analysis of internal and external feedback in self-regulated learning activities mediated by self-regulated learning tools and open learner models," *International Journal of Educational Technology in Higher Education*, vol. 17, p. 55, Dec. 2020. 2.3

[102] D. L. Butler and P. H. Winne, "Feedback and Self-Regulated Learning: A Theoretical Synthesis," *Review of Educational Research*, vol. 65, pp. 245–281, Sept. 1995. Publisher: American Educational Research Association. 2.3

[103] E. A. Locke, "Toward a theory of task motivation and incentives," *Organizational Behavior and Human Performance*, vol. 3, pp. 157–189, May 1968. 2.3.1

[104] G. P. Latham and E. A. Locke, "Increasing productivity and decreasing time limits: A field replication of Parkinson's law," *Journal of Applied Psychology*, vol. 60, no. 4, pp. 524–526, 1975. Place: US Publisher: American Psychological Association. 2.3.1

[105] A. J. Moeller, J. M. Theiler, and C. Wu, "Goal Setting and Student Achievement: A Longitudinal Study," *The Modern Language Journal*, vol. 96, no. 2, pp. 153–169, 2012. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-4781.2011.01231.x. 2.3.1, 3.1

[106] M. Morgan, "Self-monitoring of attained subgoals in private study," *Journal of Educational Psychology*, vol. 77, pp. 623–630, Dec. 1985. Publisher: American Psychological Association. 2.3.1

[107] M. Morgan, "Self-monitoring and goal setting in private study," *Contemporary Educational Psychology*, vol. 12, pp. 1–6, Jan. 1987. 2.3.1

[108] D. H. Schunk, "Goal and Self-Evaluative Influences During Children's Cognitive Skill Learning," *American Educational Research Journal*, vol. 33, pp. 359–382, June 1996. Publisher: American Educational Research Association. 2.3.1

[109] N. M. McNeil and M. W. Alibali, "Learning mathematics from procedural instruction: Externally imposed goals influence what is learned," *Journal of Educational Psychology*, vol. 92, pp. 734–744, Dec. 2000. Publisher: American Psychological Association. 2.3.1, 2.5

[110] D. Winters and G. P. Latham, "The effect of learning versus outcome goals on a simple versus a complex task," *Group & Organization Management*, vol. 21, no. 2, pp. 236–250, 1996. Place: US Publisher: Sage Publications. 2.3.2

[111] G. H. Seijts and G. P. Latham, "The effect of distal learning, outcome, and proximal goals on a moderately complex task," *Journal of Organizational Behavior*, vol. 22, no. 3, pp. 291–307, 2001. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/job.70. 2.3.2

[112] V. Aleven, B. McLaren, I. Roll, and K. Koedinger, "Toward Meta-cognitive Tutoring: A Model of Help Seeking with a Cognitive Tutor," *International Journal of Artificial Intelligence in Education*, vol. 16, pp. 101–128, Jan. 2006. 2.4

[113] K. R. Koedinger and A. Corbett, "Cognitive Tutors: Technology Bringing Learning Sciences to the Classroom," in *The Cambridge Handbook of the Learning Sciences* (R. K. Sawyer, ed.), pp. 61–78, Cambridge University Press, 1 ed., Apr. 2005. 2.4

[114] R. Azevedo, A. Witherspoon, A. Chauncey, C. Burkett, and A. Fike, "MetaTutor: A MetaCognitive tool for enhancing self-regulated learning," in *Cognitive and Metacognitive Educational Systems - Papers from the AAAI Fall Symposium, Technical Report*, pp. 14–19, Dec. 2009. 2.4

[115] R. Azevedo, R. S. Landis, R. Feyzi-Behnagh, M. Duffy, G. Trevors, J. M. Harley, F. Bouchet, J. Burlison, M. Taub, N. Pacampara, M. Yeasin, A. K. M. M. Rahman, M. I. Tanveer, and

G. Hossain, “The effectiveness of pedagogical agents’ prompting and feedback in facilitating co-adapted learning with metatutor,” in *Proceedings of the 11th international conference on Intelligent Tutoring Systems*, ITS’12, (Berlin, Heidelberg), pp. 212–221, Springer-Verlag, June 2012. 2.4

[116] G. Trevors, M. Duffy, and R. Azevedo, “Note-taking within MetaTutor: interactions between an intelligent tutoring system and prior knowledge on note-taking and learning,” *Educational Technology Research and Development*, vol. 62, pp. 507–528, Oct. 2014. 2.4

[117] S. Qiu, U. Gadiraju, and A. Bozzon, “Towards Memorable Information Retrieval,” in *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*, ICTIR ’20, (New York, NY, USA), pp. 69–76, Association for Computing Machinery, Sept. 2020. 2.5

[118] L. Nelson, C. Held, P. Pirolli, L. Hong, D. Schiano, and E. H. Chi, “With a little help from my friends: examining the impact of social annotations in sensemaking tasks,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1795–1798, New York, NY, USA: Association for Computing Machinery, Apr. 2009. 2.5

[119] M. Heilman and M. Eskenazi, “Language Learning: Challenges for Intelligent Tutoring Systems,” in *Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains*, 2006. 2.5

[120] M. Heilman, K. Collins-Thompson, J. Callan, M. Eskenazi, A. Juffs, and L. Wilson, “Personalization of Reading Passages Improves Vocabulary Acquisition,” *International Journal of Artificial Intelligence in Education*, vol. 20, pp. 73–98, Jan. 2010. Publisher: IOS Press. 2.5

[121] S. Davies, K. Butcher, and C. Stevens, “Self-Regulated Learning with Graphical Overviews: When Spatial Information Detracts from Learning,” *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 35, no. 35, 2013. 2.5

[122] W. R. Hersh, D. L. Elliot, D. H. Hickam, S. L. Wolf, and A. Molnar, “Towards new measures of information retrieval evaluation,” in *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR ’95*, (Seattle, Washington, United States), pp. 164–170, ACM Press, 1995. 2.5

[123] P.-L. Lei, C.-T. Sun, S. S. J. Lin, and T.-K. Huang, “Effect of metacognitive strategies and verbal-imagery cognitive style on biology-based video search and learning performance,” *Computers & Education*, vol. 87, pp. 326–339, Sept. 2015. 2.5

[124] L. Salmerón, P. Delgado, and L. Mason, “Using eye-movement modelling examples to improve critical reading of multiple webpages on a conflicting topic,” *Journal of Computer Assisted Learning*, vol. 36, no. 6, pp. 1038–1051, 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jcal.12458. 2.5

[125] Y. Chi, S. Han, D. He, and R. Meng, “Exploring Knowledge Learning in Collaborative Information Seeking Process,” *CEUR Workshop Proceedings*, vol. 1647, p. 5, 2016. 2.5

[126] M. J. Wilson and M. L. Wilson, “A comparison of techniques for measuring sensemaking and learning within participant-generated summaries,” *Journal of the American Society for Information Science and Technology*, vol. 64, no. 2, pp. 291–306, 2013. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.22758. 2.5

[127] M. K. Singley, "The Reification of Goal Structures in a Calculus Tutor: Effects on Problem-Solving Performance," *Interactive Learning Environments*, vol. 1, no. 2, pp. 102–23, 1990. 2.5

[128] K. R. Koedinger and J. R. Anderson, "Effective Use of Intelligent Software in High School Math Classrooms," Jan. 1993. Publisher: Carnegie Mellon University. 2.5

[129] S. Senk and Z. Usiskin, "Geometry Proof Writing: A New View of Sex Differences in Mathematics Ability," *American Journal of Education*, vol. 91, no. 2, pp. 187–201, 1983. Publisher: University of Chicago Press. 2.5

[130] N. A. Jones, H. Ross, T. Lynam, P. Perez, and A. Leitch, "Mental Models: An Interdisciplinary Synthesis of Theory and Methods," *Ecology and Society*, vol. 16, no. 1, 2011. Publisher: Resilience Alliance Inc. 2.5

[131] M. T. H. Chi, S. A. Siler, H. Jeong, T. Yamauchi, and R. G. Hausmann, "Learning from human tutoring," *Cognitive Science*, vol. 25, no. 4, pp. 471–533, 2001. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog2504_1. 2.5

[132] K. Urgo and J. Arguello, "Learning assessments in search-as-learning: A survey of prior work and opportunities for future research," *Information Processing & Management*, vol. 59, p. 102821, Mar. 2022. 3, 3.1

[133] G. J. MARCHIONINI, *Computer Enhanced Practice and Introductory Algebra*. Ph.D., Wayne State University, United States – Michigan, 1974. ISBN: 9798660671500. 3.1

[134] R. W. White, S. T. Dumais, and J. Teevan, "Characterizing the Influence of Domain Expertise on Web Search Behavior," in *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, WSDM '09, (New York, NY, USA), pp. 132–141, ACM, 2009. 3.3

[135] Y. Chi, "Examining and Supporting Laypeople's Learning in Online Health Information Seeking," in *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, CHIIR '19, (New York, NY, USA), pp. 425–428, Association for Computing Machinery, Mar. 2019. 3.3

[136] K. R. Mollan, I. M. Trumble, S. A. Reifeis, O. Ferrer, C. P. Bay, P. L. Baldoni, and M. G. Hudgens, "Precise and Accurate Power of the Rank-Sum Test for a Continuous Outcome," *Journal of biopharmaceutical statistics*, vol. 30, pp. 639–648, July 2020. 4.2

[137] K. Urgo, J. Arguello, and R. Capra, "The Effects of Learning Objectives on Searchers' Perceptions and Behaviors," in *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*, (Virtual Event Norway), pp. 77–84, ACM, Sept. 2020. 4.5, 4.7.2, 4.7.3

[138] G. Fraser, "Evaluating inclusive gender identity measures for use in quantitative psychological research," *Psychology & Sexuality*, vol. 9, pp. 343–357, Oct. 2018. Publisher: Routledge _eprint: https://doi.org/10.1080/19419899.2018.1497693. 4.7.1

[139] R. Capra, J. Arguello, H. O'Brien, Y. Li, and B. Choi, "The Effects of Manipulating Task Determinability on Search Behaviors and Outcomes," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, (New York, NY, USA), pp. 445–454, ACM, 2018. event-place: Ann Arbor, MI, USA. 4.7.2

[140] J. C. Nunnally, *Psychometric theory.* Psychometric theory, New York, NY, US: McGraw-Hill, 1967. Pages: xiii, 640. 1

[141] J. M. Cortina, "What is coefficient alpha? An examination of theory and applications," *Journal of Applied Psychology*, vol. 78, pp. 98–104, 1993. Place: US Publisher: American Psychological Association. 1

[142] A. L. Odom and L. H. Barrow, "Development and application of a two-tier diagnostic test measuring college biology students' understanding of diffusion and osmosis after a course of instruction," *Journal of Research in Science Teaching*, vol. 32, no. 1, pp. 45–61, 1995. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/tea.3660320106. 4.8

[143] R. Hake, "Relationship of Individual Student Normalized Learning Gains in Mechanics with Gender , High-School Physics , and Pretest Scores on Mathematics and Spatial Visualization," vol. 8, pp. 1–14, 2002. 4.9

[144] L. Xu, X. Zhou, and U. Gadiraju, "How Does Team Composition Affect Knowledge Gain of Users in Collaborative Web Search?," in *Proceedings of the 31st ACM Conference on Hypertext and Social Media*, HT '20, (New York, NY, USA), pp. 91–100, Association for Computing Machinery, July 2020. 4.9

[145] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977. Publisher: [Wiley, International Biometric Society]. 4.10.2, 4.12

[146] R. Brydges, J. Manzone, D. Shanks, R. Hatala, S. J. Hamstra, B. Zendejas, and D. A. Cook, "Self-regulated learning in simulation-based training: a systematic review and meta-analysis," *Medical Education*, vol. 49, no. 4, pp. 368–378, 2015. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/medu.12649. 6.1

[147] V. M. Deekens, J. A. Greene, and N. G. Lobczowski, "Monitoring and depth of strategy use in computer-based learning environments for science and history," *British Journal of Educational Psychology*, vol. 88, no. 1, pp. 63–79, 2018. _eprint: https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1111/bjep.12174. 6.1, 6.5.1, 6.5.2, 7.1

[148] J. A. Greene, N. G. Lobczowski, R. Freed, B. M. Cartiff, C. Demetriou, and A. T. Panter, "Effects of a Science of Learning Course on College Students' Learning With a Computer," *American Educational Research Journal*, vol. 57, pp. 947–978, June 2020. Publisher: American Educational Research Association. 6.1, 7.1

[149] B. M. McLaren, D. M. Adams, and R. E. Mayer, "Delayed Learning Effects with Erroneous Examples: a Study of Learning Decimals with a Web-Based Tutor," *International Journal of Artificial Intelligence in Education*, vol. 25, pp. 520–542, Dec. 2015. 6.1

[150] R. Syed and K. Collins-Thompson, "Retrieval Algorithms Optimized for Human Learning," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, (New York, NY, USA), pp. 555–564, ACM, 2017. event-place: Shinjuku, Tokyo, Japan. 6.3

[151] J. M. Harley, M. Taub, R. Azevedo, and F. Bouchet, "Let's Set Up Some Subgoals: Understanding Human-Pedagogical Agent Collaborations and Their Implications for Learning and

Prompt and Feedback Compliance," *IEEE Transactions on Learning Technologies*, vol. 11, pp. 54–66, Jan. 2018. Conference Name: IEEE Transactions on Learning Technologies. 6.5.3

[152] J. Liu and N. J. Belkin, "Personalizing information retrieval for multi-session tasks: the roles of task stage and task type," in *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, (Geneva, Switzerland), pp. 26–33, Association for Computing Machinery, July 2010. 7.3

[153] J. A. Greene, J. Robertson, and L.-J. C. Costa, "Assessing self-regulated learning using think-aloud methods," in *Handbook of self-regulation of learning and performance*, Educational psychology handbook series, pp. 313–328, New York, NY, US: Routledge/Taylor & Francis Group, 2011. A.3

[154] J. A. Greene, R. D. Plumley, C. J. Urban, M. L. Bernacki, K. M. Gates, K. A. Hogan, C. Demetriou, and A. T. Panter, "Modeling temporal self-regulatory processing in a higher education biology course," *Learning and Instruction*, vol. 72, p. 101201, Apr. 2021. A.3