The Effects of Goal-setting on Learning Outcomes and Self-Regulated Learning Processes

Kelsey Urgo
University of San Francisco
California, USA
kurgo@usfca.edu

Jaime Arguello
University of North Carolina at Chapel Hill
North Carolina, USA
jarguello@unc.edu

ABSTRACT

We present a user study ($N = 40$) that investigated the role of goal-setting on learning during search. To this end, we developed a tool called the Subgoal Manager (SM). The SM was designed to help searchers break apart a learning-oriented search task into smaller subgoals. The tool enabled participants to add, delete, and modify subgoals; take notes with respect to subgoals; and mark subgoals as completed. During the study, participants completed a single learning-oriented search task and were assigned to one of two subgoal conditions. In the Subgoals condition, participants had access to the SM; were instructed to develop at least three subgoals before the search session; and could add, delete, and modify subgoals during the search session. In the NoSubgoals condition, participants were not instructed to set subgoals and were simply provided with a text editor to take notes. We investigate the effects of the subgoal condition on: (**RQ1**) learning and retention and (**RQ2**) the extent to which participants engaged in specific self-regulated learning (SRL) processes during the search session. Our results found two important trends. First, participants in the Subgoals condition had better learning outcomes, especially with respect to retention. Second, based on a qualitative analysis of participants' search sessions, participants in the Subgoals condition engaged in more self-regulated learning (SRL) processes. Combined, our results suggest that goal-setting improves learning during search by encouraging and supporting greater engagement with SRL processes.

CCS CONCEPTS

• **Information systems → Users and interactive retrieval**.

KEYWORDS

Search as learning, learning outcomes, self-regulated learning

ACM Reference Format:
Kelsey Urgo and Jaime Arguello. 2024. The Effects of Goal-setting on Learning Outcomes and Self-Regulated Learning Processes. In *Proceedings of the 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '24), March 10–14, 2024, Sheffield, United Kingdom.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3627508.3638348

CHIIR '24, March 10–14, 2024, Sheffield, United Kingdom

ACM ISBN 979-8-4007-0434-5/24/03
https://doi.org/10.1145/3627508.3638348

1 INTRODUCTION

Existing search systems are effective in helping users complete simple search tasks (e.g., fact-finding or navigational tasks). However, they provide less support for users completing complex tasks that involve *learning*. To bridge this gap, *search-as-learning* studies have investigated different tools to support learning during search. Studies have considered tools that enable searchers to annotate documents [18, 59, 64], visualizations that communicate a searcher's coverage of subtopics during the search process [12], and retrieval algorithms that help searchers learn new vocabulary [70].

Our research in this paper examines the role of goal-setting on learning during search. To this end, we developed a simple tool called the Subgoal Manager (SM) (Figure 1). The SM was designed to help searchers deconstruct a learning-oriented search task into smaller subgoals. The tool enables searchers to add, modify, and delete subgoals; take notes with respect to specific subgoals; and mark subgoals as complete. Features of the tool were designed to help searchers monitor their progress toward their subgoals. Each subgoal has its own text editor for searchers to take notes with respect to the subgoal. Additionally, checking a "subgoal complete" checkbox collapses the subgoal's text editor and turns the subgoal a darker shade of gray to visually distinguish it from incomplete subgoals. Prior work has found that subgoals are more achievable when they include precise action, content, standard to measure success, and timeframe [47]. Therefore, tooltips on the interface remind searchers to generate subgoals with these ideal characteristics.

Two bodies of literature suggest that effective goal-setting improves learning. First, education research has underscored the important role of goal-setting in self-regulated learning (SRL) [69]. SRL is an active and reflective process in which a learner monitors and controls their own learning. While several models of SRL have been proposed [8, 56, 82, 86, 87], the Winne & Hadwin (W&H) model [82] highlights goal-setting as its own distinct phase of SRL. Additionally, goals are important to effective SRL as they provide standards for learners to monitor and control their progress, as well as change strategies when needed. Prior studies have found that learners that effectively engage in SRL processes from the W&H model have better learning outcomes [14, 24]. Second, goal-setting research has found that goals help people: (1) improve their understanding of a task; (2) activate prior knowledge and skills; (3) remain focused on task-relevant activities; and (4) increase persistence [36, 37, 42].

In this paper, we report on a study with 40 participants. During the study, participants completed a single learning-oriented search task—learn everything you can about the concepts of diffusion and osmosis. The study adopted a between-subjects design and participants were assigned to one of two *subgoal conditions*. In the Subgoals condition, participants were provided with the Subgoal

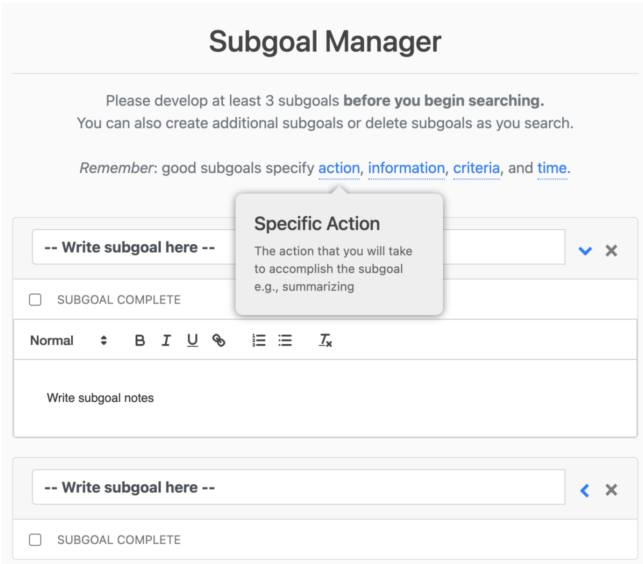


Figure 1: The Subgoal Manager

Manager, instructed to set subgoals with ideal characteristics, and asked to develop at least three subgoals before searching. Participants could add, modify, and delete subgoals during the search task. In the NoSubgoals condition, participants did not have access to the Subgoal Manager and were not explicitly instructed to set subgoals before nor during the task. Instead, participants were provided with a Text Editor tool (similar to a Google Doc) to take notes during the search task.

We investigate two main research questions:

- **RQ1:** What are the effects of the subgoal condition on learning and knowledge retention?
- **RQ2:** What are the effects of the subgoal condition on the frequency of specific SRL processes?

To investigate **RQ1**, participants completed two different types of assessments: a closed-ended assessment and an open-ended assessment. The closed-ended assessment consisted of the Osmosis and Diffusion Conceptual Assessment (ODCA) [17]. The multiple-choice ODCA was developed by experts and targets common misconceptions about diffusion and osmosis. Participants completed the ODCA before the search task (to capture prior knowledge), immediately after the search task (to measure learning), and one week later (to measure knowledge retention). While multiple-choice tests are easy to grade, they may not capture everything that someone learned about a topic. Therefore, participants also completed an open-ended assessment that asked them to describe "everything you learned during the search task". Participants completed the open-ended assessment immediately after the search task (to measure learning) and one week later (to measure knowledge retention).

To investigate **RQ2**, the study used a think-aloud protocol. That is, participants were asked to narrate their thoughts during the search task. Participants' think-aloud comments and screen activities were recorded and analyzed using qualitative techniques. Based on observed behaviors, we developed a hierarchical taxonomy of macro- and micro-SRL processes. Our set of SRL processes is an extension of a framework developed by Greene et al. [22, 23], which was rooted in the W&H model of SRL. Our taxonomy includes 3 macro-SRL processes and 36 micro-SRL processes.

Our research in this paper builds on our own prior work [75]. In a previous crowdsourced study, we found that participants who used the Subgoal manager and set their own goals had slightly better learning outcomes. The present study extends our prior work in several important ways. First, we measured learning using two types of assessments. This enabled us to get a more complete and accurate picture of the effects of goal-setting on learning during search. Second, we measured knowledge retention by having participants complete both learning assessments one week after the search task. Finally, and most importantly, we are the first to investigate the role of goal-setting on the extent to which searchers engage in different SRL processes during a learning-oriented search task. Our RQ2 results suggest that goal-setting improves learning during search partly *because* it supports more effective SRL.

2 BACKGROUND

Our research builds on three areas of prior work: (1) search-as-learning, (2) self-regulated learning, and (3) goal-setting.

Search-as-Learning: Learning is an important goal of search [5, 16, 62]. The search-as-learning community was established to better support learning during search [3, 10]. Search-as-learning work has studied factors that impact learning during search: (1) characteristics of the learning objective [9, 29, 32, 83]; (2) characteristics of the individual [53, 55, 63, 79]; and (3) characteristics of the system [15, 18, 31, 58, 64, 71]. Research has also explored search behaviors that predict learning during search [1, 7, 11, 19, 39, 46, 54, 85].

Particularly relevant to our study, prior search-as-learning studies have explored novel interfaces and tools to better support learning during search. Kammerer et al. [31] investigated the effects of a search system that enabled participants to filter the search results using social tags. Participants had better learning outcomes with the experimental system. Câmara et al. [12] explored the effects of an experimental search interface that presented participants with their coverage of subtopics across the search session. With the experimental system, participants explored more subtopics *superficially* and, ultimately, did not have better learning outcomes. Freund et al. [18] investigated the effects of two interface features on learning. One manipulation involved displaying articles in plain text versus HTML with distracting elements. A second manipulation involved giving participants access to an auxiliary tool to annotate articles. Without the tool, participants had better learning outcomes in the plain text condition. With the tool, participants had similar learning outcomes in both display conditions. Roy et al. [64] explored the effects of a system that enabled participants to highlight text and make notes during the search session. Individually, note-taking and highlighting had positive effects on learning. However, participants who used *both* tools did not have better learning outcomes possibly due to cognitive overload. Syed et al. [72] developed a system that prompted participants to answer automatically generated questions about passages read during the search session. The experimental system had positive effects on learning. Qiu et al. [59] investigated the effects of two factors on learning: (1) search vs. chatbot and

(2) note-taking vs. no note-taking. Participants had better learning outcomes with the search interface and when asked to take notes.

Measuring learning is critically important in search-as-learning work. Prior studies have used a wide range of learning assessment methods [73]. Assessment methods can be classified as closed- or open-ended. Closed-ended assessments have included: (1) multiple-choice tests [13, 18, 26, 27, 30, 70, 77]; (2) true-or-false tests [18, 19, 30, 52, 59, 85]; and (3) short-answer tests [1, 11–13, 28, 63, 64]. Open-ended assessments have asked participants to: (1) list key phrases and facts [7, 31]; (2) generate visual representations of a topic [40]; (3) list pro and con arguments toward a proposition [15]; and (4) summarize what was learned [1, 11, 13, 30, 38, 39, 53–55, 65, 79]. Studies have scored open-ended responses by: (1) counting pertinent key phrases or facts [1, 7, 11, 31, 54, 79]; (2) counting pertinent pro and con arguments [15]; and (3) counting statements that demonstrate critical thinking [1, 11, 39, 53, 54, 65].

To investigate **RQ1**, our study used *both* a closed-ended and open-ended assessment. Our closed-ended assessment consisted of a well-established instrument called the Osmosis and Diffusion Conceptual Assessment (ODCA) [17]. The ODCA is a multiple-choice test that was designed to target common misconceptions about diffusion and osmosis. The ODCA has mostly been used in education research. For example, past research has used the ODCA to explore common misconceptions of biology students [60] and to measure prior knowledge during an evaluation of an immersive educational environment [61].

Self-Regulated Learning: Learning sciences research has investigated the important role of self-regulated learning (SRL) in improving learning outcomes [8, 68, 69, 80, 88]. SRL is an active and reflective process in which a learner monitors and controls their own learning [21, 80]. Although several models of SRL exist, the Winne & Hadwin (W&H) model [82] is widely accepted and has been used in many empirical studies [4, 6, 20, 21, 33, 66, 76]. Additionally, and particularly important to our study, the W&H model emphasizes the critical function of goal-setting in learning. The W&H model consists of four iterative phases—(1) *Task Definition*; (2) *Goals & Plans*; (3) *Strategies & Tactics*; and (4) *Adaptations*. Throughout each phase, learners engage in metacognitive monitoring (e.g., tracking progress toward goals) and control (e.g., implementing different strategies to achieve goals).

Capturing the extent to which learners engage in SRL processes can be difficult. Prior work has used two methods—(1) asking about participants' perceptions [57, 67, 78] and (2) using a think-aloud protocol [22–24]. Both methods have benefits and drawbacks. Self-report questionnaires are easy to implement. However, perceptions of SRL engagement may not align with *actual* SRL engagement [81]. Think-aloud protocols allow researchers to manually categorize think-aloud comments and observable behaviors into specific SRL processes. This method captures SRL processes as they occur. However, it involves significant effort. In prior work, we explored the effects of goal-setting on *perceptions* of SRL engagement [75]. To our knowledge, we are the first to capture *actual* SRL engagement based on think-aloud comments and observable behavior during a search session. Similarly, we are the first to study the effects of goal-setting on *actual* SRL engagement during search.

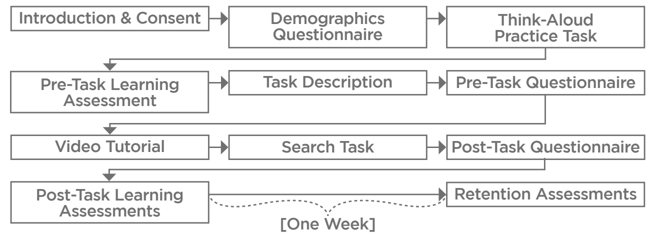


Figure 2: Study Protocol

Goal-Setting: Goals play an important function in SRL. Research has found that goals—(1) prompt a learner to consider their understanding of the task [37]; (2) focus a learner's attention toward task planning and task-relevant strategies [42]; and (3) provide standards for monitoring progress [47]. Prior work has shown that goals are likely to improve learning outcomes [35, 49–51, 69]. This work has largely focused on students setting goals in a traditional classroom setting. Our study is distinct in that it focuses on searchers learning *independently* in the context of a *search environment*. Prior work has also identified ideal goal characteristics that make them more achievable. In particular, goals should be—(1) difficult; (2) specific; (3) proximal (short-term); (4) learning-oriented (versus performance-oriented); and (5) self-set [41, 43–45]. With respect to specificity, research has found that ideal goals include a specific action (i.e., cognitive process), content (i.e., material), standard (i.e., criteria to measure progress and success) and allotted timeframe [47]. In our study, participants in the Subgoals condition watched a training video describing these ideal characteristics.

3 METHODS

To investigate **RQ1** & **RQ2**, we conducted a study with 40 participants ($F = 28$, $M = 12$). Participants were undergraduate students at our university and were 18 to 22 years old ($M = 19.93$, $S.D. = 1.07$). Participants completed a search task that involved the concepts of diffusion and osmosis. Therefore, participants were required to *not* be biology nor chemistry majors. Participants were asked about the highest level of biology course taken. Participants reported taking a graduate-level course (1), undergraduate-level course (24), high school-level course (14), and none (1). The study used a between-subjects design. Participants were assigned to one of two conditions: NoSubgoals or Subgoals (20 participants per condition). The study was approved by our university's Institutional Review Board (IRB).

3.1 Study Protocol

Figure 2 illustrates the study protocol. The study took place over two sessions conducted one week apart. Both sessions were conducted remotely using the Zoom videoconferencing platform. The first session proceeded as follows. First, participants watched a video describing the study, signed a consent form, and completed a demographics questionnaire. Second, participants completed a practice think-aloud search task. Third, participants completed the multiple-choice ODCA to measure their prior knowledge of diffusion and osmosis. Fourth, participants were given the main search task and were asked to read it aloud. Then, after completing a pre-task questionnaire, participants watched a video introducing the tools to be used during the search task (i.e., the search system and the Text Editor or Subgoal Manager). Participants then completed the main search task, which was limited to 40 minutes. During the search task, the moderator had their microphone and camera turned off and only unmuted themselves to remind participants to think aloud as needed. After the search task, participants completed a post-task questionnaire. Finally, to measure learning during the search task, participants completed the multiple-choice ODCA a second time and an open-ended assessment that asked them to "describe everything you learned during the search task". Participants did not have access to their notes while completing these assessments. Participants were given $US30 after the first study session. The second study session took place one week later. During the second study session, to measure knowledge retention, participants completed the multiple-choice ODCA a third time and the open-ended assessment a second time. Participants were given $US10 after the second study session. All participants completed both sessions. Given our focus in this paper, we do not report on results from our pre- and post-task questionnaires. Briefly, pre-task perceptions were not significantly different between groups and participants reported significantly greater engagement with SRL planning and monitoring progress in the Subgoals condition.

3.2 Experimental Conditions

The study had two experimental conditions.

Subgoals: In this condition, participants were provided with the Subgoal Manager (SM) with blank subgoals (Figure 1). Before the search task, participants watched a video that described the functionality of the SM and the search system. The video instructed participants on how to add, delete, and modify subgoals; take notes with respect to subgoals; and mark subgoals as completed. As mentioned in Section 2, prior research has found that ideal subgoals specify an action, content, standard, and timeframe. Therefore, the video also described these four ideal subgoal characteristics. Participants were shown an example of an ideal subgoal: "Spend 10 minutes [time] identifying [action] three paintings [standard] that demonstrate the main characteristics of surrealism [content]." The video pointed to the part of the subgoal associated with each ideal characteristic. Finally, the video instructed participants to develop at least three subgoals before starting the search task.

NoSubgoals: In this condition, participants were provided with the Text Editor tool (similar to a Google Doc) to take notes during the search task. Before the search task, participants watched a video that described the functionality of the Text Editor and the search system. Participants were not instructed to set subgoals before nor during the search task.

3.3 Search System

To gather information, participants interacted with a custom-built search system implemented using the Bing Web Search API. The system returned four types of results in different tabs: webpages, images, news articles, and videos. Given a query, the system returned 50 results per type. Web, news, and video results were displayed as a ranked list with 10 results per page. All 50 images results were displayed in a grid layout. We configured the Bing API to return results for the US-EN market and had safe-mode turned on. Three buttons along the top of the search interface allowed participants to: (1) (re-)open the Subgoal Manager or Text Editor, depending on the subgoal condition; (2) (re-)open the task description; and (3) indicate when they finished the search task.

3.4 Search Task

Each participant completed the following learning-oriented search task, which included a scenario to contextualize the task and a main objective:

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.

Objective: 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.

We used this task for several reasons. First, it is a conceptual learning task. We wanted to study the role of goal-setting during a *complex* learning task. Prior research has found that conceptual learning during search is more complex than factual or procedural learning [74]. Second, prior work has found that goal-setting improves motivation and persistence [42]. In this respect, we wanted the task to allow participants to set subgoals with different standards. Our chosen task was fairly open-ended, allowing participants to set subgoals with different standards toward the overall task to "learn everything you can." Finally, and most importantly, the topic of the task allowed us to use the well-established ODCA to measure learning and retention.

3.5 ODCA Learning Assessment

To measure learning and retention, participants completed the multiple-choice ODCA before the search task, immediately after, and one week later. The ODCA was developed by Fisher et al. [17] and includes 18 questions about diffusion and osmosis. The questions are organized in pairs. Each pair contains a knowledge question and a reasoning question. The knowledge question is designed to assess the student's comprehension of specific concepts and processes related to diffusion and osmosis. The reasoning question is designed to assess the student's justification for their answer to the corresponding knowledge question. In other words, the knowledge question focuses on "what?" and the reasoning question focuses on "why?". For example, one knowledge question is: "All cells are: (a) semipermeable or (b) permeable". In this case, the reasoning question asks: "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; or (d) allow some substances to pass through, but not others."

The ODCA was used for two main reasons. First, it targets common misconceptions that even biology students have about the concepts of diffusion and osmosis [17]. Second, the ODCA is a valid and reliable instrument [17]. To establish face validity, the ODCA was developed with the help of expert biology faculty. Additionally, a panel of students participated in semi-structured interviews in which 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. In terms of reliability, ODCA items have been found to have high internal consistency across student cohorts [17].

To measuring learning, we used participants' pre- and post-task scores on the ODCA to compute *normalized gain:*

$$\text{Normalized Gain} = \frac{(\text{PostScore} - \text{PreScore})}{(1 - \text{PreScore})}, \quad (1)$$

where PreScore and PostScore are the percentage of correct answers in the pre- and post-task ODCA, respectively. Similarly, to measure retention, we used the same normalization. That is, we used Equation 1 but replaced PostScore with RetScore—the percentage of correct answers in the ODCA retention test. This type of normalization is common in education research [25] and search-as-learning studies [19, 84, 85]. Normalized gain accounts for participants' prior knowledge based on their pre-task scores. It essentially answers the question: "Of the percentage a participant could have gained, what percentage did they actually gain?" On rare occasions, the normalized gain can be a negative value (i.e., PostScore < PreScore). This can happen due to participants guessing correctly on the pre-test and not the post-test.

3.6 Open-Ended Learning Assessment

Multiple-choice tests such as the ODCA are easy to grade but may not capture everything that someone learned. For this reason, participants also completed open-ended assessments immediately after the search task and one week later. This assessment asked participants to "describe everything you learned during the search task." Participants were provided with a text box to enter their response. Next, we describe how open-ended responses were scored.

Our goal was to analyze each open-ended response as a set of correct and incorrect statements. A statement is defined as a logical unit that is either entirely true or entirely false. A sentence can contain multiple statements. For example, "Diffusion and osmosis are forms of passive transport." contains two statements: (1) "Diffusion is a form of passive transport" and (2) "Osmosis is a form of passive transport." Our decision to analyze open-ended responses as sets of (in)correct statements enabled us to give participants partial credit for sentences containing both true and false statements.

In total, we identified 374 unique statements across all 80 open-ended responses. These unique statements were labeled as true or false by two biology professors. Both experts reviewed statements jointly and resolved any disagreements through discussion. We asked the experts to provide a brief justification for statements labeled as false. Statements were labeled as false for several reasons. First, some statements were labeled as false because the opposite is true. For example, the statement "facilitated diffusion requires energy" was labeled as false because "facilitated diffusion does not require energy". Second, some statements were labeled as false because a word or concept was misremembered. For example, "hypotropic indicates low concentration" was marked as false because "hypotonic [not hypotropic] indicates low concentration". Third, statements with an overstated claim or incorrect premise were labeled as false. For example, "particles naturally want to move from areas of high-to-low concentration" was labeled as false because "particles don't 'want' anything". Finally, some statements were labeled as false because the participant confused two related concepts. For example, "active diffusion involves ATP" was marked as false because "active transport [not diffusion] involves ATP". Ultimately, 62 of 374 (17%) unique statements were labeled as false.

Our analysis of open-ended responses required two manual processes: (1) splitting sentences into statements and (2) mapping statements that are semantically equivalent. The second manual process was required in order to identify the set of 374 unique statements (labeled as true or false as described above) and mapping the statements in each response to one of these 374 unique statements. Next, we describe how these two manual processes were validated.

Splitting Sentences into Statements: To validate this process, two authors on the paper (A1 and A2) analyzed all open-ended responses from four participants (10% of the data). Each author worked independently to identify all unique statements in each response. To measure agreement, we used the Jaccard coefficient, which measures the intersection divided by the union of unique statements identified by A1 and A2. The Jaccard coefficient (averaged across responses) was 0.83. Given this high level of agreement, A1 split all sentences into statements across all responses.

Semantically Equivalent Statements: This manual process was validated as follows. First, A1 identified the set of 374 unique statements across all responses. Then, A1 and A2 independently analyzed all statements associated with open-ended responses from four participants (10% of the data). A1 and A2 independently assigned each statement to one of the 374 unique statements identified by A1. The Cohen's Kappa agreement between A1 and A2 was $\kappa = 0.982$, which is considered "almost perfect" agreement [34]. Given this high level of agreement, A1 was responsible mapping all statements to one of the 374 unique statements.

Measuring Learning & Retention: After all the steps above, we were able to treat each open-ended response as a set of true and false statements. To measure learning, we computed the percentage of true statements in the post-task open-ended response. To measure retention, we computed two measures. First, we computed the percentage of true statements in the retention open-ended response. Second, we computed the percentage of true statements retained between the post-task and retention open-ended assessments.

3.7 Qualitative Coding of SRL Processes

In **RQ2**, we investigate the effects of the subgoal condition on the extent to which participants engaged in specific SRL processes during the search task. To address **RQ2**, we conducted a qualitative analysis of participants' search sessions by leveraging their recorded think-aloud comments and screen activities, which included search, reading and note-taking activities. SRL processes were coded using the SRL Coding Guide shown in Tables 1-3. The SRL Coding Guide was adapted from Greene et al. [22, 23]. The coding guide consists of three macro-SRL processes: *Planning*, *Strategy Use*, and *Monitoring*. Each macro-SRL process contains 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 1-3 provide the definition and an example of each micro-SRL process. Seven micro-SRL processes are new and were created in order to capture SRL behaviors observed in our study. These new micro-SRL processes are likely to be observed in other search studies and are marked with (*) in Tables 1-3.

Our qualitative analysis of participants' search sessions involved two authors on the paper (A1 and A2). First, A1 segmented 100% of all search sessions into so-called "codable units". Codable units were identified based on behaviors and/or think-aloud comments

Table 1: Macro-SRL Process: Planning

Micro-SRL Process	Description	Example
*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."
Recycles 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 2: Macro-SRL Process: Strategy Use

Micro-SRL Process	Description	Example
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 3: Macro-SRL Process: Monitoring

Micro-SRL Process	Description	Example
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**	Assesses 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."

that indicate engagement with a specific micro-SRL process. For example, a behavior such as highlighting a sentence while taking notes indicates engagement with the micro-SRL process *Emphasizing Notes*. Similarly, a think-aloud comment such as "now I'll focus on how diffusion is different from osmosis" indicates engagement with the micro-SRL process *Subgoals*.

After segmenting search sessions into codable units, the data was coded in three rounds. During each round of coding, each codable unit was assigned to a single 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 also assigned to a single macro-SRL process. For example, if a codable unit was assigned to the micro-SRL process of *Emphasizing Notes* it was also assigned to the macro-SRL process of *Strategy Use*. During the first round of coding, A1 and A2 independently coded search sessions from 4 participants. After this initial round, A1 and A2 met to discuss disagreements and modify the SRL Coding Guide. In the second round of coding, A1 and A2 again independently coded search sessions from an additional 4 participants. The Cohen's Kappa agreement was $\kappa = 0.872$ for micro-SRL processes and $\kappa = 0.944$ for macro-SRL processes, which are both considered "almost perfect" agreement [34]. Then, A1 recoded the 4 participants from the first round and an additional 24 participants for a total of 32 participants. After this, A1 and A2 independently coded search sessions from 2 participants. This was done to ensure that A1 had not drifted from the code definitions validated during the second round of coding. The Cohen's Kappa agreement was $\kappa = 0.919$ for micro-SRL processes and $\kappa = 0.944$ for macro-SRL processes, which are considered "almost perfect" agreement [34]. Finally, A1 coded the remaining 6 participants. Our analysis was based on codes assigned by A1.

4 RESULTS

In this section, we report on results for **RQ1** & **RQ2**. Our analyses involved comparisons between participants assigned to the Subgoals and NoSubgoals conditions. Most outcome measures were not normally distributed. Therefore, we tested for statistically significant differences between groups (i.e., Subgoals vs. NoSubgoals) using non-parametric Mann-Whitney U tests. To conserve space, we only show figures for outcome measures with statistically significant differences between groups. All figures correspond to box plots. In the text, we use M_S and M_N to denote the median for the Subgoals condition and NoSubgoals condition, respectively.

4.1 Differences in Prior Knowledge

Our study used a between-subjects design. Therefore, it was important to verify that participants in both conditions had similar levels of prior knowledge about diffusion and osmosis. Prior knowledge was measured using the percentage of correct answers in the pre-task ODCA assessment. There were no statistically significant differences between groups. In fact, both groups had the same median score ($M_S = 0.528$, $M_N = 0.528$).[1]

4.2 RQ1: Effects on Learning and Retention

RQ1 investigated the effects of the subgoal condition on learning and knowledge retention.

[1]While there was a difference in standard deviation between groups ($\sigma_S = 0.14$, $\sigma_N = 0.16$), the median pre-task ODCA score was identical between groups.

Effects on Learning: To measure learning, participants completed the ODCA before and after the search task. We used participants' scores on these assessments to compute normalized gain (Equation 1). Additionally, participants completed the open-ended assessment after the search task. Open-ended assessments were scored based on the percentage of true statements included. We did not find statistically significant differences between groups for either measure. However, on average, participants in the Subgoals condition had higher normalized gains based on their pre- and post-task ODCA scores ($M_S = 0.41$ vs. $M_N = 0.22$) and included a *slightly* greater percentage of true statements in their post-task open-ended responses ($M_S = 0.90$ vs. $M_N = 0.86$).

Effects on Knowledge Retention: To measure knowledge retention, participants completed the ODCA and open-ended assessment one week after the search task. Again, we used participants' pre-task and retention ODCA scores to compute normalized gain. Additionally, we used participants' open-ended responses to compute two measures: (1) the percentage of correct statements included in the retention open-ended response and (2) the percentage of correct statements retained between the post-task and retention open-ended response.

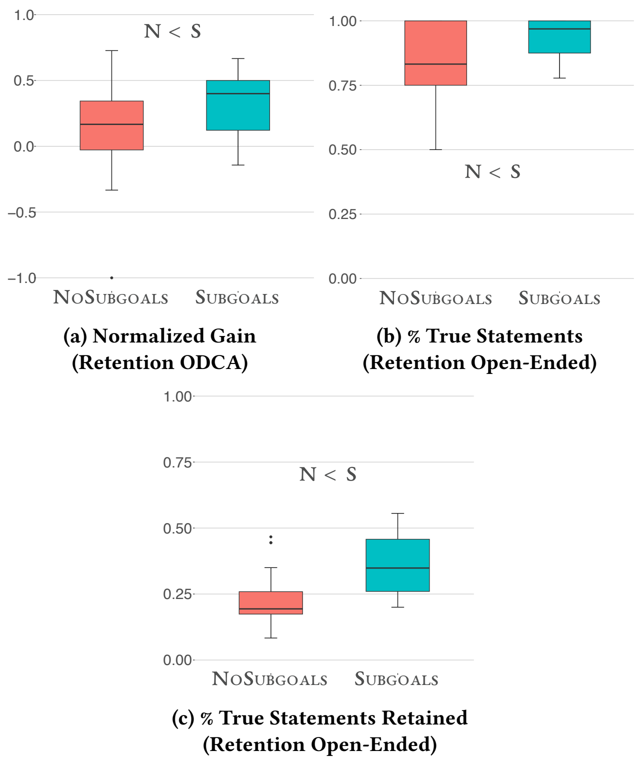
Figure 3a shows differences in normalized gain on the retention ODCA between conditions. Differences in normalized gain on the retention ODCA were statistically significant ($U = 127$, $p < .05$). On average, participants in the Subgoals condition had higher normalized gains in the retention ODCA ($M_S = 0.40$ vs. $M_N = 0.17$).

Figures 3b and 3c show differences in the percentage of true statements on the open-ended retention assessment and the percent of true statements retained across open-ended assessments between conditions. There were statistically significant differences detected between conditions in terms of the percentage of true statements included in the open-ended retention assessment ($U = 118$, $p < .05$) and the percentage of true statements retained across open-ended assessments ($U = 77$, $p < .001$). On average, participants in the Subgoals condition included a greater percentage of true statements in the open-ended retention assessment ($M_S = 0.97$ vs. $M_N = 0.83$) and retained a greater percentage of true statements across open-ended assessments ($M_S = 0.35$ vs. $M_N = 0.19$).

4.3 RQ2: Effects on SRL Engagement

RQ2 investigated the effects of the subgoal condition on the extent to which participants engaged in different macro- and micro-SRL processes during the search task. First, we report on frequency with which participants engaged in specific macro-SRL processes. Second, we report on the *diversity* of micro-SRL processes that participants engaged in within each macro-SRL process. Third, we report on the frequency with which participants engaged in specific micro-SRL processes.

Frequency of Macro-SRL Processes: Figures 4a-4c show the frequencies with which participants engaged in the macro-SRL processes of *Planning*, *Monitoring*, and *Strategy Use*. Differences between groups were found to be statistically significant for *Planning* ($U = 21.5$, $p < .0001$) and *Monitoring* ($U = 114$, $p < .05$). On average, participants in the Subgoals condition engaged in more *Planning* ($M_S = 21$ vs. $M_N = 8$) and *Monitoring* ($M_S = 28.5$ vs. $M_N = 20.5$). Differences between groups were marginally significant for *Strategy Use* ($U = 141.5$, $p = .06$). On average, participants



(a) Normalized Gain (Retention ODCA)

(b) % True Statements (Retention Open-Ended)

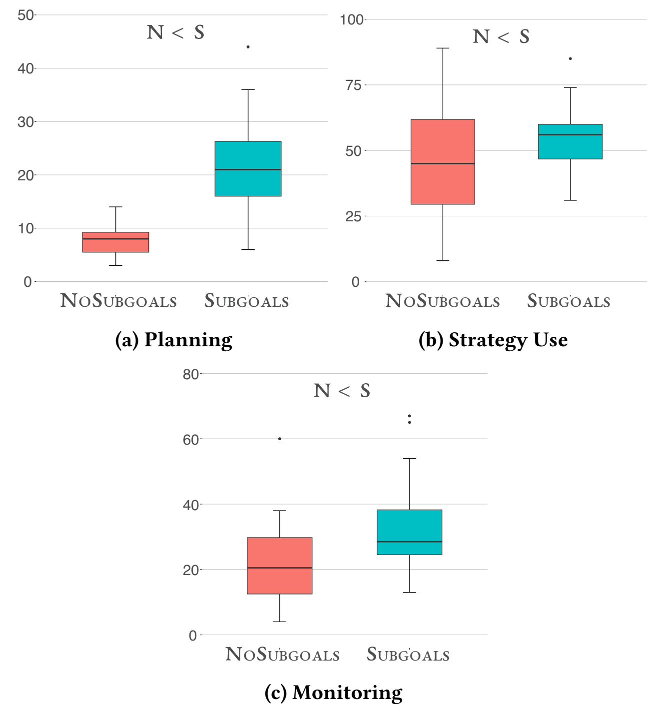
(c) % True Statements Retained (Retention Open-Ended)

Figure 3: Learning based on Retention Assessments

in the Subgoals condition engaged in more *Strategy Use* ($M_S = 56$ vs. $M_N = 45$).

Diversity of Micro-SRL Processes: Figures 5a-5b show the number of *distinct* micro-SRL processes that participants engaged in within the macro-SRL processes of *Planning* and *Strategy Use*. Differences between groups were found to be statistically significant for *Planning* ($U = 71$, $p < .001$) and *marginally* significant for *Strategy Use* ($U = 000$, $p = .06$). On average, participants in the Subgoals condition engaged in a more diverse set of micro-SRL processes related to *Planning* ($M_S = 5$ vs. $M_N = 3$) and *Strategy Use* ($M_S = 9$ vs. $M_N = 8$). Differences between groups were not statistically significant for *Monitoring*.

Frequency of Micro-SRL Processes: Table 4 shows the frequencies with which participants engaged in specific micro-SRL processes. For brevity, with one exception, we only include those micro-SRL processes with statistically significant differences between groups ($p < .05$). Within the macro-SRL process of *Planning*, participants in the Subgoals condition engaged in signficantly more: (1) *Subgoals*; (2) *Recycles Goal in Working Memory*; (3) *Revisits Previous Subgoals*; and (4) *Modifies Subgoals*. Within the macro-SRL process of *Strategy Use*, participants in the Subgoals condition engaged in significantly more *Prior Knowledge Activation*. Participants also engaged in more *Comparing & Contrasting*. However, these differences were marginally significant ($p = .05$). Finally, within the macro-SRL process of *Monitoring*, participants in the Subgoals condition engaged in significantly more: (1) *Monitor Progress Toward Subgoals*; (2) *Monitor Subgoal Quality*; (3) *Expectation of Adequacy of Content*; and (4) *Time Monitoring*.

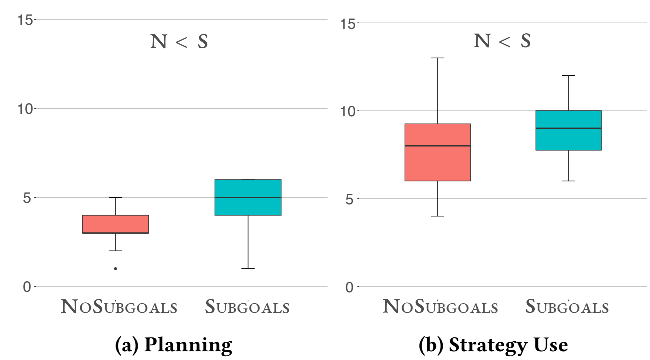


(a) Planning

(b) Strategy Use

(c) Monitoring

Figure 4: Frequency of Macro-SRL Processes



(a) Planning

(b) Strategy Use

Figure 5: Diversity of micro-SRL processes

5 DISCUSSION

RQ1: Effects on Learning and Retention: RQ1 investigated the influence of the subgoal condition on learning during search. Learning was measured using two types of assessments: the multiple-choice 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 task. The open-ended assessment was administered immediately after and one week after the search task.

Our **RQ1** results found that participants in the Subgoals condition had better learning outcomes, particularly with respect to knowledge retention. Participants in the Subgoals condition had significantly higher normalized gains based on their ODCA scores

Table 4: Frequency of Micro-SRL Processes with Signficant Differences. Median (Min.,Max.)

Micro-SRL Process	NoSugboals	Sugboals
Planning		
Subgoals	4.5 (0,9)	7.5 (4,17)
Recycles Goal in Working Memory	0 (0,3)	1 (0,2)
Revisits Previous Subgoal	0.5 (0,3)	6 (0,16)
Modifies Subgoals	0 (0,1)	1 (0,10)
Strategy Use		
Prior Knowledge Activation	0.5 (0,3)	3 (0,6)
Comparing & Contrasting	0 (0,5)	1.5 (0,9)
Monitoring		
Monitor Progress Toward Subgoals	2 (0,3)	4.5 (1,15)
Monitor Subgoal Quality	0 (0,0)	0 (0,2)
Expectation of Adequacy of Content	3.5 (0,23)	8 (0,18)
Time Monitoring	0 (0,3)	1 (0,9)

before the search task and one week later. Responses to the open-ended assessments found similar trends. First, participants in the Subgoals condition included a significantly greater percentage of correct statements in their responses to the open-ended assessment one week after the search task. Second, participants in the Subgoals condition retained a significantly greater percentage of correct statements between their responses to the open-ended assessment immediately after the search task and one week later.

Our **RQ1** results extend our own prior work. In a previous paper [75], we reported on a crowdsourced study that also used the Subgoal Manager to investigate the effects of goal-setting on learning during search. That study involved three experimental conditions. Two conditions were similar to our NoSubgoals and Subgoals conditions. A third condition (AssignedSubgoals) involved the Subgoal Manager with pre-populated subgoals. In that study, participants in the Subgoals (versus NoSubgoals) condition scored *slightly* higher on the ODCA taken immediately after the search task. In the current study, we found a similar trend. However, our **RQ1** results also show that participants in the Subgoals (versus NoSubgoals) condition were better able to *retain* what they learned. This is an important contribution of the current study.

Our **RQ1** results have three important implications. The first implication relates to tools to support learning during search. The second and third implications relate to methodological recommendations for future search-as-learning studies.

First and foremost, our results suggest that goal-setting improves learning during search. To encourage and support goal-setting, we designed the Subgoal Manager. The Subgoal Manager included simple features for participants to set goals and monitor their progress toward their goals (e.g., take notes with respect to subgoals and mark subgoals as completed). Our results suggest that search systems should incorporate such tools to support learning. Importantly, prior studies have found that note-taking can improve learning during search [18, 59, 64]. In our study, participants took notes in both conditions. Therefore, our results suggest that goal-setting and *goal-specific* note-taking can provide additional benefits.

Second, while we did not find statistically significant differences between groups based on assessments administered immediately after the search task, we did find significant differences based on assessments administered one week later. Studies outside of search-as-learning have found similar trends. For example, McLaren et al. [48] and Adams et al. [2] evaluated a tutoring system that asked students to correct an erroneous solution to a problem versus simply solve the problem. While both groups performed similarly on an assessment administered immediately after the learning session, participants in the "erronous solution" condition performed better on a delayed assessment. The authors argued that the "erroneous solution" condition resulted in deeper, longer-lasting learning. With few exceptions [59, 72], most search-as-learning studies have only measured knowledge gains immediately after the search task. Our results underscore the importance of also measuring retention.

Third, our study used two different types of assessments—the multiple-choice ODCA and an open-ended assessment. The ODCA provided a simple mechanism for comparing between groups. However, multiple-choice tests may not capture everything that someone learned. Therefore, the open-ended assessment enabled us to capture knowledge gains not captured by the ODCA. In retrospect, including the open-ended assessment was a good decision. In total, participants included 374 unique statements in their open-ended responses. Therefore, it is highly likely that participants learned things that were not targeted by the ODCA. The use of multiple assessments provided stronger evidence that participants in the Subgoals condition retained more of what they learned. By administering multiple assessment types, researchers in search-as-learning may also strengthen their findings in similar ways.

RQ2: Effects on SRL Engagement: RQ2 investigated the influence of the subgoal condition on the extent to which participants engaged in different macro- and micro-SRL processes. To this end, participants' recorded think-aloud comments and actions were coded into SRL processes. Our results found that participants in the Subgoals condition had higher levels of engagement with all 3 macro-SRL processes and 10 (out of 36) micro-SRL processes.

Participants in the Subgoals condition engaged in more *Planning*. Within *Planning*, participants engaged in more: (1) *Subgoals*; (2) *Recycles Goal in Working Memory*; (3) *Revisits Previous Subgoals*; and (4) *Modifies Subgoals.* Participants in the Subgoals condition were asked to set at least three subgoals before the search task. Therefore, it is not surprising that they set more subgoals and subsequently modified specific subgoals during the task. It is important to note, however, that participants in the NoSubgoals condition *did* develop subgoals, just at a lower rate (M_N = 4.5 vs. M_S = 7.5). This is important because it shows that while subgoals are beneficial, searchers may not naturally develop them. Therefore, tools should explicitly prompt searchers to develop subgoals and allow them to modify subgoals during the search session.

Participants in the Subgoals condition were also more likely to return to a previously set subgoal and to recycle the current subgoal in working memory (e.g., restate it verbally as a way to stay focused). Features of the Subgoal Manager may have influenced these behaviors. For example, being able to see their subgoals may have prompted participants to shift between subgoals. Many of these shifts happened when participants encountered information that was pertinent to a subgoal other than the current one. Similarly, participants in the Subgoals condition were asked to create subgoals with a specific action, content, standard, and timeframe. Therefore, participants in the Subgoals condition may have had clearer, more well-defined aims, which they were able to articulate verbally. Based on these results, tools should keep subgoals visible and encourage searchers to set subgoals with ideal characteristics.

Participants in the Subgoals condition engaged in more *Strategy Use*. Within *Strategy Use*, participants engaged in more: (1) *Comparing & Contrasting* and (2) *Prior Knowledge Activation.* As previously mentioned, participants sometimes encountered information that was relevant to a subgoal they had previously worked on. This often prompted participants to switch subgoals and to compare/contrast the information with their notes on the previous subgoal. For example, while learning about the definition of diffusion, one participant encountered a definition of osmosis that differed from the one in their notes. This prompted them to revisit their definition of osmosis (i.e., revisit previous subgoal) and compare/contrast it with the one encountered. This trend suggests that the Subgoal Manager may have enabled participants to compartmentalize information more effectively, perhaps because they had clearer objectives and notes explicitly associated with those objectives. This may have enabled participants to recognize relevant information about a subgoal other than the current one.

Participants in the Subgoals condition also engaged in a higher frequency of *Prior Knowledge Activation* (*PKA*). This is an interesting result given that *PKA* was not *directly* supported in the Subgoals condition. 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 goal-setting as a way to activate task-relevant knowledge [47, 80].

Participants in the Subgoals condition engaged in more *Monitoring*. Within, *Monitoring*, participants engaged in more: (1) *Monitor Progress Toward Subgoals*; (2) *Monitor Subgoal Quality*; (3) *Expectation of Adequacy of Content*; and (4) *Time Monitoring*. Features of the Subgoal Manager directly supported participants in monitoring their progress toward specific subgoals. The interface included a "subgoal complete" button and displayed the amount of notes associated with each subgoal, which may have prompted participants to monitor their progress toward a subgoal. Participants in the Subgoals condition were instructed to set subgoals with ideal characteristics. Additionally, tooltips on the Subgoal Manager reminded participants about these ideal subgoal characteristics. This may have prompted participants to more closely monitor the quality of their subgoals. Among these ideal subgoal characteristics, participants were instructed to specify a timeframe for each subgoal. It is less clear why participants in the Subgoals condition engaged in more *Expectation of Adequacy of Content.* One possible explanation is that participants were asked to develop subgoals with a clearly defined content and standard. This may have prompted participants to be more selective when deciding whether to engage with information resources more deeply.

Our **RQ2** results also found that participants in the Subgoals condition engaged in more diverse micro-level processes related to *Planning* and *Strategy Use.* It is somewhat expected that participants in the Subgoals condition engaged in more diverse micro-SRL processes related to *Planning*. Participants in the Subgoals condition developed subgoals and modified subgoals during the search process. Additionally, having subgoals visible may have prompted them to recycle the current subgoal in working memory and revisit a previous subgoal. It is interesting that participants in the Subgoals condition also engaged in more diverse micro-SRL processes related to *Strategy Use.* Participants in the Subgoals condition were asked to develop subgoals with specific standards. It is possible that standards prompted participants to enact different strategies in order to meet the standards. For example, participants often developed subgoals about gathering a specific number of definitions, examples, or similarities/differences between diffusion and osmosis. This often prompted participants to take inventory of their prior knowledge (i.e., *Prior Knowledge Activation*) and (re-)read their notes to see if they had met the subgoal (i.e., *Reading Notes*). Perhaps the greater diversity of micro-SRL processes contributed to participants having better learning outcomes in the Subgoals condition.

6 CONCLUSION

In this paper, we have presented results from a study that investigated the effects of goal-setting on: (**RQ1**) learning and knowledge retention and (**RQ2**) levels of engagement with specific SRL processes during the search session.

Learning & SRL Engagement: Our **RQ1** results found that participants in the Subgoals condition had better learning outcomes. Specifically, they were better able to retain knowledge gained during the search session. Our **RQ2** results found that participants in the Subgoals condition had higher levels of engagement with SRL processes related to *Planning*, *Monitoring*, and *Strategy Use.* Combined, our results shed light onto *why* participants had better learning outcomes in the Subgoals condition. It is likely that the higher levels of SRL engagement contributed to deeper, longer-lasting learning. Prior studies in education have also found a positive relation between SRL engagement and learning outcomes [14, 24]. Our study extends this prior work in three ways. First, these studies had students learn about a subject by interacting with documents in a hypermedia environment. In contrast, our study focused on interactive search using a web search engine. Second, our study involved a system manipulation (i.e., access to the Subgoal Manager). Third, our study measured learning immediately after the search session *and* one week later.

Future Work: Several open questions remain. First, in the Subgoals condition, participants had access to the Subgoal Manager, were instructed on how to set good subgoals, and were asked to set at least three subgoals before starting the search task. Future work should investigate the impact of these decisions. For example, how important is it to coach searchers on how to set good subgoals? Similarly, how important is it to set subgoals *before* starting the search task? Second, future work should investigate whether specific SRL processes are more *critical* at certain stages of a search task. Finally, future work could extend the Subgoal Manager by adding new features. Examples include a feature that automatically detects suboptimal subgoals (e.g., lacking standards to measure success) or a feature that highlights information that is relevant to a specific subgoal (i.e., not necessarily the current one).

Acknowledgements: This work was supported by NSF grant IIS-2106334. Any opinions, findings, conclusions, and recommendations expressed in this paper are the authors' and do not necessarily reflect those of the sponsor.

REFERENCES

[1] Mustafa Abualsaud. 2017. *Learning Factors and Determining Document-level Satisfaction In Search-as-Learning.* Master's thesis. University of Waterloo, Waterloo, Ontario, Canada.

[2] Deanne M. Adams, Bruce M. McLaren, Kelley Durkin, Richard E. Mayer, Bethany Rittle-Johnson, Seiji Isotani, and Martin van Velsen. 2014. Using erroneous examples to improve mathematics learning with a web-based tutoring system. *Computers in Human Behavior* 36 (2014), 401–411.

[3] James Allan, Bruce Croft, Alistair Moffat, and Mark Sanderson. 2012. Frontiers, challenges, and opportunities for information retrieval: Report from SWIRL 2012 the second strategic workshop on information retrieval in Lorne. *ACM SIGIR Forum* 46, 1 (May 2012), 2. https://doi.org/10.1145/2215676.2215678

[4] Roger Azevedo, Susan Ragan, Jennifer G. Cromley, and Stacy Pritchett. 2002. *Do Different Goal-Setting Conditions Facilitate Students' Ability to Regulate Their Learning of Complex Science Topics with RiverWeb?* https://eric.ed.gov/?id=ED482509

[5] Peter Bailey, L Chen, and S Grosenick. 2012. User task understanding: a web search engine perspective. *NII Shonan workshop* (Oct. 2012), 17. http://research.nii.ac.jp/ntcir/shonan/Bailey-NIIShonan-UserTaskUnderstanding.pdf

[6] Maria Bannert, Melanie Hildebrand, and Christoph Mengelkamp. 2009. Effects of a metacognitive support device in learning environments. *Computers in Human Behavior* 25, 4 (July 2009), 829–835. https://doi.org/10.1016/j.chb.2008.07.002

[7] Nilavra Bhattacharya and Jacek Gwizdka. 2019. 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.* ACM, Glasgow Scotland UK, 63–71. https://doi.org/10.1145/3295750.3298926

[8] Monique Boekaerts, Moshe Zeidner, Paul R. Pintrich, and Paul R Pintrich. 1999. *Handbook of Self-Regulation.* Elsevier Science & Technology, San Diego, UNITED STATES. http://ebookcentral.proquest.com/lib/unc/detail.action?docID=300645

[9] Robert Capra, Jaime Arguello, Anita Crescenzi, and Emily Vardell. 2015. 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).* ACM, New York, NY, USA, 23–32. https://doi.org/10.1145/2766462.2767741

[10] Kevyn Collins-Thompson, Preben Hansen, and Claudia Hauff. 2017. Search as Learning (Dagstuhl Seminar 17092). (2017). https://doi.org/10.4230/dagrep.7.2.135

[11] Kevyn Collins-Thompson, Soo Young Rieh, Carl C. Haynes, and Rohail Syed. 2016. 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).* ACM, New York, NY, USA, 163–172. https://doi.org/10.1145/2854946.2854972

[12] Arthur Câmara, Nirmal Roy, David Maxwell, and Claudia Hauff. 2021. Searching to Learn with Instructional Scaffolding. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval (CHIIR '21).* Association for Computing Machinery, New York, NY, USA, 209–218. https://doi.org/10.1145/3406522.3446012

[13] Sarah Davies, Kirsten Butcher, and Corey Stevens. 2013. Self-Regulated Learning with Graphical Overviews: When Spatial Information Detracts from Learning. *Proceedings of the Annual Meeting of the Cognitive Science Society* 35, 35 (2013). https://escholarship.org/uc/item/5kp8b0n9

[14] Victor M. Deekens, Jeffrey A. Greene, and Nikki G. Lobczowski. 2018. Monitoring and depth of strategy use in computer-based learning environments for science and history. *British Journal of Educational Psychology* 88, 1 (2018), 63–79. https://doi.org/10.1111/bjep.12174 _eprint: https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1111/bjep.12174.

[15] Diego Demaree, Halszka Jarodzka, Saskia Brand-Gruwel, and Yvonne Kammerer. 2020. 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).* Association for Computing Machinery, New York, NY, USA, 373–377. https://doi.org/10.1145/3343413.3378000

[16] Carsten Eickhoff, Jaime Teevan, Ryen White, and Susan Dumais. 2014. 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.* ACM Press, New York, New York, USA, 223–232. https://doi.org/10.1145/2556195.2556217

[17] Kathleen M. Fisher, Kathy S. Williams, and Jennifer Evarts Lineback. 2011. Osmosis and Diffusion Conceptual Assessment. *CBE—Life Sciences Education* 10, 4 (Dec. 2011), 418–429. https://doi.org/10.1187/cbe.11-04-0038 Publisher: American Society for Cell Biology (lse).

[18] Luanne Freund, Rick Kopak, and Heather O'Brien. 2016. The effects of textual environment on reading comprehension: Implications for searching as learning. *Journal of Information Science* 42, 1 (Feb. 2016), 79–93. https://doi.org/10.1177/0165551515614472 Publisher: SAGE Publications Ltd.

[19] Ujwal Gadiraju, Ran Yu, Stefan Dietze, and Peter Holtz. 2018. 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).* ACM, New York, NY, USA, 2–11. https://doi.org/10.1145/3176349.3176381

[20] Inga Glogger, Rolf Schwonke, Lars Holzäpfel, Matthias Nückles, and Alexander Renkl. 2012. Learning strategies assessed by journal writing: Prediction of learning outcomes by quantity, quality, and combinations of learning strategies. *Journal of Educational Psychology* 104, 2 (2012), 452–468. https://doi.org/10.1037/a0026683 Place: US Publisher: American Psychological Association.

[21] Jeffrey Alan Greene and Roger Azevedo. 2007. A Theoretical Review of Winne and Hadwin's Model of Self-Regulated Learning: New Perspectives and Directions. *Review of Educational Research* 77, 3 (Sept. 2007), 334–372. https://doi.org/10.3102/003465430303953 Publisher: American Educational Research Association.

[22] Jeffrey Alan Greene, Cheryl Mason Bolick, William P. Jackson, Alfred M. Caprino, Christopher Oswald, and Megan McVea. 2015. Domain-specificity of self-regulated learning processing in science and history. *Contemporary Educational Psychology* 42 (July 2015), 111–128. https://doi.org/10.1016/j.cedpsych.2015.06.001

[23] Jeffrey Alan Greene, Dana Z. Copeland, Victor M. Deekens, and Seung B. Yu. 2018. Beyond knowledge: Examining digital literacy's role in the acquisition of understanding in science. *Computers & Education* 117 (Feb. 2018), 141–159. https://doi.org/10.1016/j.compedu.2017.10.003

[24] Jeffrey A. Greene, Nikki G. Lobczowski, Rebekah Freed, Brian M. Cartiff, Cynthia Demetriou, and A. T. Panter. 2020. Effects of a Science of Learning Course on College Students' Learning With a Computer. *American Educational Research Journal* 57, 3 (June 2020), 947–978. https://doi.org/10.3102/0002831219865221 Publisher: American Educational Research Association.

[25] R. Hake. 2002. Relationship of Individual Student Normalized Learning Gains in Mechanics with Gender , High-School Physics , and Pretest Scores on Mathematics and Spatial Visualization, Vol. 8. 1–14. https://www.semanticscholar.org/paper/Relationship-of-Individual-Student-Normalized-Gains-Hake/ab557de0fdafe5def057a795c25264e74ac0e332

[26] Michael Heilman, Kevyn Collins-Thompson, Jamie Callan, Maxine Eskenazi, Alan Juffs, and Lois Wilson. 2010. Personalization of Reading Passages Improves Vocabulary Acquisition. *International Journal of Artificial Intelligence in Education* 20, 1 (Jan. 2010), 73–98. https://doi.org/10.3233/JAI-2010-0003 Publisher: IOS Press.

[27] Michael Heilman and Maxine Eskenazi. 2006. Language Learning: Challenges for Intelligent Tutoring Systems. In *Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains.*

[28] William R. Hersh, Diane L. Elliot, David H. Hickam, Stephanie L. Wolf, and Anna Molnar. 1995. 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.* ACM Press, Seattle, Washington, United States, 164–170. https://doi.org/10.1145/215206.215355

[29] Bernard J. Jansen, Danielle Booth, and Brian Smith. 2009. Using the taxonomy of cognitive learning to model online searching. *Information Processing & Management* 45, 6 (Nov. 2009), 643–663. https://doi.org/10.1016/j.ipm.2009.05.004

[30] Rishita Kalyani and Ujwal Gadiraju. 2019. Understanding User Search Behavior Across Varying Cognitive Levels. In *Proceedings of the 30th ACM Conference on Hypertext and Social Media (HT '19).* Association for Computing Machinery, New York, NY, USA, 123–132. https://doi.org/10.1145/3342220.3343643

[31] Yvonne Kammerer, Rowan Nairn, Peter Pirolli, and Ed H. Chi. 2009. 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).* Association for Computing Machinery, Boston, MA, USA, 625–634. https://doi.org/10.1145/1518701.1518797

[32] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan-ching Wu. 2015. 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).* ACM, New York, NY, USA, 101–110. https://doi.org/10.1145/2808194.2809465

[33] Saskia Kistner, Katrin Rakoczy, Barbara Otto, Charlotte Dignath-van Ewijk, Gerhard Büttner, and Eckhard Klieme. 2010. Promotion of self-regulated learning in classrooms: investigating frequency, quality, and consequences for student performance. *Metacognition and Learning* 5, 2 (Aug. 2010), 157–171. https://doi.org/10.1007/s11409-010-9055-3

[34] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (1977), 159–174. https://doi.org/10.2307/2529310 Publisher: [Wiley, International Biometric Society].

[35] Gary P. Latham and Travor C. Brown. 2006. The Effect of Learning vs. Outcome Goals on Self-Efficacy, Satisfaction and Performance in an MBA Program. *Applied Psychology: An International Review* 55, 4 (2006), 606–623. https://doi.org/10.1111/j.1464-0597.2006.00246.x Place: United Kingdom Publisher: Blackwell Publishing.

[36] Gary P. Latham, Terence R. Mitchell, and Dennis L. Dossett. 1978. Importance of participative goal setting and anticipated rewards on goal difficulty and job performance. *Journal of Applied Psychology* 63, 2 (1978), 163–171. https://doi.org/10.1037/0021-9010.63.2.163 Place: US Publisher: American Psychological Association.

[37] Gary P. Latham and Lise M. Saari. 1979. Application of social-learning theory to training supervisors through behavioral modeling. *Journal of Applied Psychology* 64, 3 (1979), 239–246. https://doi.org/10.1037/0021-9010.64.3.239 Place: US Publisher: American Psychological Association.

[38] Pei-Lan Lei, Chuen-Tsai Sun, Sunny S. J. Lin, and Tsung-Kuan Huang. 2015. Effect of metacognitive strategies and verbal-imagery cognitive style on biology-based video search and learning performance. *Computers & Education* 87 (Sept. 2015), 326–339. https://doi.org/10.1016/j.compedu.2015.07.004

[39] Chang Liu and Xiaoxuan Song. 2018. How do Information Source Selection Strategies Influence Users' Learning Outcomes'. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18)*. Association for Computing Machinery, New York, NY, USA, 257–260. https://doi.org/10.1145/3176349.3176876

[40] Hanrui Liu, Chang Liu, and Nicholas J. Belkin. 2019. Investigation of users' knowledge change process in learning-related search tasks. *Proceedings of the Association for Information Science and Technology* 56, 1 (2019), 166–175. https://doi.org/10.1002/pra2.63

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

[42] Edwin A. Locke and Gary P. Latham. 2002. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist* 57, 9 (Sept. 2002), 705–717. https://doi.org/10.1037/0003-066X.57.9.705 Publisher: American Psychological Association.

[43] Edwin A. Locke and Gary P. Latham. 2006. New Directions in Goal-Setting Theory. *Current Directions in Psychological Science* 15, 5 (Oct. 2006), 265–268. https://doi.org/10.1111/j.1467-8721.2006.00449.x Publisher: SAGE Publications Inc.

[44] Edwin A. Locke and Gary P. Latham. 2012. *New Developments in Goal Setting and Task Performance.* Routledge, London, UNITED KINGDOM. http://ebookcentral.proquest.com/lib/unc/detail.action?docID=1104793

[45] Edwin A. Locke and Gary P. Latham. 2019. The development of goal setting theory: A half century retrospective. *Motivation Science* 5, 2 (June 2019), 93–105. https://doi.org/10.1037/mot0000127 Publisher: Educational Publishing Foundation.

[46] Yihan Lu and I-Han Hsiao. 2017. Personalized Information Seeking Assistant (PiSA): from programming information seeking to learning. *Information Retrieval Journal* 20, 5 (Oct. 2017), 433–455. https://doi.org/10.1007/s10791-017-9305-y

[47] Lindsay McCardle, Elizabeth A. Webster, Adrianna Haffey, and Allyson F. Hadwin. 2017. Examining students' self-set goals for self-regulated learning: Goal properties and patterns. *Studies in Higher Education* 42, 11 (Nov. 2017), 2153–2169. https://doi.org/10.1080/03075079.2015.1135117 Publisher: Routledge _eprint: https://doi.org/10.1080/03075079.2015.1135117.

[48] Bruce M. McLaren, Deanne M. Adams, and Richard E. Mayer. 2015. Delayed Learning Effects with Erroneous Examples: a Study of Learning Decimals with a Web-Based Tutor. *International Journal of Artificial Intelligence in Education* 25, 4 (Dec. 2015), 520–542. https://doi.org/10.1007/s40593-015-0064-x

[49] Aleidine J. Moeller, Janine M. Theiler, and Chaorong Wu. 2012. Goal Setting and Student Achievement: A Longitudinal Study. *The Modern Language Journal* 96, 2 (2012), 153–169. https://doi.org/10.1111/j.1540-4781.2011.01231.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-4781.2011.01231.x.

[50] Mark Morgan. 1985. Self-monitoring of attained subgoals in private study. *Journal of Educational Psychology* 77, 6 (Dec. 1985), 623–630. https://doi.org/10.1037/0022-0663.77.6.623 Publisher: American Psychological Association.

[51] Mark Morgan. 1987. Self-monitoring and goal setting in private study. *Contemporary Educational Psychology* 12, 1 (Jan. 1987), 1–6. https://doi.org/10.1016/S0361-476X(87)80033-4

[52] Les Nelson, Christoph Held, Peter Pirolli, Lichan Hong, Diane Schiano, and Ed H. Chi. 2009. 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*. Association for Computing Machinery, New York, NY, USA, 1795–1798. https://doi.org/10.1145/1518701.1518977

[53] Heather L. O'Brien, Andrea Kampen, Amelia W. Cole, and Kathleen Brennan. 2020. The Role of Domain Knowledge in Search as Learning. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval (CHIIR '20)*. Association for Computing Machinery, Vancouver BC, Canada, 313–317. https://doi.org/10.1145/3343413.3377989

[54] Srishti Palani, Zijian Ding, Stephen MacNeil, and Steven P. Dow. 2021. 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)*. Association for Computing Machinery, New York, NY, USA, 325–329. https://doi.org/10.1145/3406522.3446046

[55] Georg Pardi, Johannes von Hoyer, Peter Holtz, and Yvonne Kammerer. 2020. 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*. ACM, Vancouver BC Canada, 378–382. https://doi.org/10.1145/3343413.3378001

[56] Paul R. Pintrich. 2000. Chapter 14 - The Role of Goal Orientation in Self-Regulated Learning. In *Handbook of Self-Regulation*, Monique Boekaerts, Paul R. Pintrich, and Moshe Zeidner (Eds.). Academic Press, San Diego, 451–502. https://doi.org/10.1016/B978-012109890-2/50043-3

[57] Paul R. Pintrich and Elisabeth V. de Groot. 1990. Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology* 82, 1 (1990), 33–40. https://doi.org/10.1037/0022-0663.82.1.33 Place: US Publisher: American Psychological Association.

[58] Sihang Qiu, Alessandro Bozzon, and Ujwal Gadiraju. 2020. Conversational interfaces for search as learning. In *CIKMW2020: Proceeding of the CIKM 2020 Workshops*. CEUR. https://research.tudelft.nl/en/publications/conversational-interfaces-for-search-as-learning

[59] Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. 2020. Towards Memorable Information Retrieval. In *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval (ICTIR '20)*. Association for Computing Machinery, New York, NY, USA, 69–76. https://doi.org/10.1145/3409256.3409830

[60] Nicole B. Reinke, Mary Kynn, and Ann L. Parkinson. 2019. Conceptual understanding of osmosis and diffusion by Australian first-year biology students. *International Journal of Innovation in Science and Mathematics Education* 27, 9 (2019). https://doi.org/10.30722/IJISME.27.09.002 Number: 9.

[61] Nicole B. Reinke, Mary Kynn, and Ann L. Parkinson. 2021. Immersive 3D Experience of Osmosis Improves Learning Outcomes of First-Year Cell Biology Students. *CBE—Life Sciences Education* 20, 1 (March 2021), ar1. https://doi.org/10.1187/cbe.19-11-0254

[62] Soo Young Rieh, Kevyn Collins-Thompson, Preben Hansen, and Hye-Jung Lee. 2016. Towards searching as a learning process: A review of current perspectives and future directions. *Journal of Information Science* 42, 1 (Feb. 2016), 19–34. https://doi.org/10.1177/0165551515615841

[63] Nirmal Roy, Felipe Moraes, and Claudia Hauff. 2020. Exploring Users' Learning Gains within Search Sessions. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval (CHIIR '20)*. Association for Computing Machinery, Vancouver BC, Canada, 432–436. https://doi.org/10.1145/3343413.3378012

[64] Nirmal Roy, Manuel Valle Torre, Ujwal Gadiraju, David Maxwell, and Claudia Hauff. 2021. 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)*. Association for Computing Machinery, New York, NY, USA, 229–238. https://doi.org/10.1145/3406522.3446025

[65] Ladislao Salmerón, Pablo Delgado, and Lucia Mason. 2020. Using eye-movement modelling examples to improve critical reading of multiple webpages on a conflicting topic. *Journal of Computer Assisted Learning* 36, 6 (2020), 1038–1051. https://doi.org/10.1111/jcal.12458 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jcal.12458.

[66] Radhika Santhanam, Sharath Sasidharan, and Jane Webster. 2008. Using Self-Regulatory Learning to Enhance E-Learning-Based Information Technology Training. *Information Systems Research* 19, 1 (March 2008), 26–47. https://doi.org/10.1287/isre.1070.0141 Publisher: INFORMS.

[67] Gregory Schraw and Rayne Sperling Dennison. 1994. Assessing Metacognitive Awareness. *Contemporary Educational Psychology* 19, 4 (Oct. 1994), 460–475. https://doi.org/10.1006/ceps.1994.1033

[68] Dale H. Schunk and Carl W. Swartz. 1993. Goals and Progress Feedback: Effects on Self-Efficacy and Writing Achievement. *Contemporary Educational Psychology* 18, 3 (July 1993), 337–354. https://doi.org/10.1006/ceps.1993.1024

[69] Traci Sitzmann and Katherine Ely. 2011. 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* 137, 3 (2011), 421–442. https://doi.org/10.1037/a0022777 Place: US Publisher: American Psychological Association.

[70] Rohail Syed and Kevyn Collins-Thompson. 2017. Optimizing search results for human learning goals. *Information Retrieval Journal* 20, 5 (Oct. 2017), 506–523. https://doi.org/10.1007/s10791-017-9303-0

[71] Rohail Syed, Kevyn Collins-Thompson, Paul N. Bennett, Mengqiu Teng, Shane Williams, Dr. Wendy W. Tay, and Shamsi Iqbal. 2020. Improving Learning Outcomes with Gaze Tracking and Automatic Question Generation. In *Proceedings of The Web Conference 2020 (WWW '20)*. Association for Computing Machinery, New York, NY, USA, 1693–1703. https://doi.org/10.1145/3366423.3380240

[72] Rohail Syed, Kevyn Collins-Thompson, Paul N Bennett, Mengqiu Teng, Shane Williams, Dr Wendy W Tay, and Shamsi Iqbal. 2020. Improving learning outcomes with gaze tracking and automatic question generation. In *Proceedings of The Web Conference 2020*. 1693–1703.

[73] Kelsey Urgo and Jaime Arguello. 2022. Learning assessments in search-as-learning: A survey of prior work and opportunities for future research. *Information Processing & Management* 59, 2 (March 2022), 102821. https://doi.org/10.1016/j.ipm.2021.102821

[74] Kelsey Urgo and Jaime Arguello. 2022. Understanding the "Pathway" Towards a Searcher's Learning Objective. *ACM Transactions on Information Systems* 40, 4 (Jan. 2022), 77:1–77:42. https://doi.org/10.1145/3495222

[75] Kelsey Urgo and Jaime Arguello. 2023. Goal-setting in support of learning during search: An exploration of learning outcomes and searcher perceptions. *Information Processing & Management* 60, 2 (March 2023), 103158. https://doi.org/10.1016/j.ipm.2022.103158

[76] Gerard van den Boom, Fred Paas, and Jeroen J. G. van Merriënboer. 2007. Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes. *Learning and Instruction* 17, 5 (Oct. 2007), 532–548. https://doi.org/10.1016/j.learninstruc.2007.09.003

[77] Nino Weingart and Carsten Eickhoff. 2016. Retrieval Techniques for Contextual Learning. *SAL @ SIGIR* (2016), 5.

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

[79] Teena Willoughby, S. Alexandria Anderson, Eileen Wood, Julie Mueller, and Craig Ross. 2009. Fast searching for information on the Internet to use in a learning context: The impact of domain knowledge. *Computers & Education* 52, 3 (April 2009), 640–648. https://doi.org/10.1016/j.compedu.2008.11.009

[80] Philip H. Winne and Allyson F. Hadwin. 1998. Studying as self-regulated engagement in learning. In *Metacognition in educational theory and practice*.

[81] Philip H. Winne, Dianne Jamieson-Noel, and Krista Muis. 2002. Methodological issues and advances in researching tactics, strategies, and self-regulated learning. In *New directions in measures and methods* (1. ed ed.), Paul R. Pintrich and Martin L. Maehr (Eds.). Number 12 in Advances in motivation and achievement. JAI, An Imprint of Elsevier Science, Amsterdam.

[82] Philip H. Winne and Nancy E. Perry. 2000. Measuring self-regulated learning. In *Handbook of self-regulation*. Academic Press, San Diego, CA, US, 531–566. https://doi.org/10.1016/B978-012109890-2/50045-7

[83] Wan-Ching Wu, Diane Kelly, Ashlee Edwards, and Jaime Arguello. 2012. 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)*. ACM, New York, NY, USA, 254–257. https://doi.org/10.1145/2362724.2362768 event-place: Nijmegen, The Netherlands.

[84] Luyan Xu, Xuan Zhou, and Ujwal Gadiraju. 2020. 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)*. Association for Computing Machinery, New York, NY, USA, 91–100. https://doi.org/10.1145/3372923.3404784

[85] Ran Yu, Ujwal Gadiraju, Peter Holtz, Markus Rokicki, Philipp Kemkes, and Stefan Dietze. 2018. Predicting User Knowledge Gain in Informational Search Sessions. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*. ACM, New York, NY, USA, 75–84. https://doi.org/10.1145/3209978.3210064

[86] Barry J. Zimmerman. 2000. Chapter 2 - Attaining Self-Regulation: A Social Cognitive Perspective. In *Handbook of Self-Regulation*, Monique Boekaerts, Paul R. Pintrich, and Moshe Zeidner (Eds.). Academic Press, San Diego, 13–39. https://doi.org/10.1016/B978-012109890-2/50031-7

[87] Barry J. Zimmerman. 2002. Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice* 41, 2 (May 2002), 64–70. https://doi.org/10.1207/s15430421tip4102_2

[88] Barry J. Zimmerman and Dale H. Schunk. 2011. *Handbook of self-regulation of learning and performance.* Routledge/Taylor & Francis Group, New York, NY, US. Pages: xiv, 484.