



# Supporting Metacognition during Exploratory Search with the OrgBox

Anita Crescenzi, Austin R. Ward, Yuan Li, Rob Capra

University of North Carolina at Chapel Hill

amcc@unc.edu, austinrw@unc.edu, yuanli@live.unc.edu, rcapra@unc.edu

## ABSTRACT

Current search systems provide effective support to users engaged in fact-finding and look-up oriented tasks. However, they provide relatively little support for users engaged in exploratory search tasks that involve cognitive and metacognitive activities such as learning, synthesis, planning, and reflection. We conducted a within-subject user study (N=24) that investigated the effects of a novel knowledge organization tool called the OrgBox, designed to assist users with organizing and synthesizing information, and metacognitive activities. The OrgBox included features to allow users to drag-drop information they found through search into "boxes" that could be created, labelled, and re-arranged. Study participants completed two exploratory search tasks, one with the OrgBox, and one with the OrgDoc, a baseline tool that included features of a rich-text editor (e.g., formatting, bullets) for taking notes.

In this paper, we present results from our study comparing the OrgBox and OrgDoc tools. Specifically, we investigate if there were differences in participants' (1) search interactions, (2) saving and organizing behaviors (e.g., amount of information, structure of notes), (3) perceptions of the tasks, tool usability, and quality of their task outputs, and (4) perceptions of how the tools provided support for cognitive and metacognitive activities involved in the task. Our results show that when using the OrgBox tool, participants created more grouping sections in their notes and saved more text. In terms of metacognitive support, participants perceived the OrgBox tool to provide significantly higher levels of support for three types of metacognitive activity (monitoring/tracking, evaluation, and planning) without changing their perceptions of the task difficulty.

## CCS CONCEPTS

• **Information systems** → **Users and interactive retrieval**; • **Human-centered computing** → *Human computer interaction (HCI)*;

## KEYWORDS

exploratory search; metacognition; knowledge organization tool; OrgBox

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

SIGIR '21, July 11–15, 2021, Virtual Event, Canada

© 2021 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.

ACM ISBN 978-1-4503-8037-9/21/07...\$15.00

<https://doi.org/10.1145/3404835.3462955>

## ACM Reference Format:

Anita Crescenzi, Austin R. Ward, Yuan Li, Rob Capra. 2021. Supporting Metacognition during Exploratory Search with the OrgBox. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21)*, July 11–15, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3404835.3462955>

## 1 INTRODUCTION

By their nature, exploratory tasks involve cognitively demanding activities including discovery, learning, interpretation, and comparison [27, 41]. In addition, successful exploratory searches often require users to engage in *metacognitive* activities such as monitoring progress toward (sub-)goals, assessing ones' current state of knowledge, and planning a strategy to move forward [7, 26].

Researchers in the field of education have shown that specific training and tools can encourage students to engage in metacognitive behaviors and can improve learning outcomes [8, 29, 44]. Related work in information retrieval (IR) by the "search as learning" community is investigating how learning during search occurs and how it can be supported by search interfaces and tools [10, 16, 18]. Research on information-seeking has long recognized the importance of metacognitive activities during search. For example, metacognitive activities play important roles in choosing strategies to use during search [5, 6], sensemaking [14, 33, 36], synthesizing information [27], and the use of idea tactics as cognitive strategies intended to improve a searchers' thinking [3]. Furthermore, recent IR workshops have called for more research on integrated tools to help support users engaged in complex search tasks [13, 23].

To explore these issues, our lab developed an integrated search assistance tool, the OrgBox, to help users by providing integrated support for search and information saving, organizing, and synthesizing [40]. We designed the OrgBox based on our prior work exploring how users take paper-and-pencil notes during an exploratory search [12]. The OrgBox includes features to allow users to drag-and-drop information they find during searches into "boxes" and "items" that can be created, labelled, and re-arranged on a canvas. We designed the OrgBox tool with the goal of providing users additional support for organization, synthesis, and metacognitive activities (e.g., planning, evaluation, monitoring). It was designed to support activities such as comparing items, identifying information gaps, and revising structure by rearranging the positions of boxes and items. A recent paper found evidence that an initial version of the OrgBox tool provided searchers with metacognitive benefits compared to a baseline bookmarking tool [9].

In this paper, we investigate the potential benefits of the OrgBox tool compared against a baseline tool called the OrgDoc that allowed participants to save and structure information in a rich-text

editor (e.g., text formatting, indentation, bullets). We wanted to understand if the OrgBox would provide more metacognitive benefits than this type of commonly used note-taking tool. We conducted an exploratory, within-subjects user study (N=24). Participants completed two exploratory search tasks, one with the OrgBox, and one with the OrgDoc. Collectively, we refer to the OrgBox and OrgDoc as the OrgTools.

Specifically, we investigated four research questions:

**RQ1.** Are there differences in search interactions between tasks completed with the OrgBox vs. the OrgDoc tools?

**RQ2.** Are there differences in information saving and organization behaviors between the OrgBox and OrgDoc tools (e.g., the amount of information saved and the number of major structures)?

**RQ3.** Are there differences in participants' perceptions of the (a) usability of the OrgTool or search system, (b) the quality of the set of notes they produced for the task, and (c) the topics and tasks between the OrgBox and the OrgDoc?

**RQ4.** Are there differences in participants' perceptions of how the OrgTools provided support for metacognitive activities during the task?

To address these questions, we analyzed logged interactions with the search and OrgBox systems (RQ1, RQ2), structures used in the sets of notes derived from content analysis (RQ2), and perceptions measured in pre- and post-task questionnaires (RQ3, RQ4).

## 2 RELATED WORK

### 2.1 Support for Exploratory Search

Exploratory search is an iterative, multi-tactical process, that requires users to explore an information space to fulfill an open-ended information need [41]. Information search for exploratory tasks is complex and involves active engagement from users during the search process [27]. Specifically, Marchionini [27] describes that in addition to looking up information, exploratory searchers often need to learn and investigate the found information which involves higher-level cognitive and metacognitive activities (e.g., analysis, synthesis, and evaluation).

**2.1.1 Tools integrated with search systems.** Researchers working at the intersections of HCI and IR have developed tools that are integrated with search systems to help users find, re-find, and keep track of search progress. For example, Morris et al. [31] developed the *SearchBar* tool to help users resume searches for complex tasks and to re-find information they had seen previously. The *SearchBar* captured and stored queries issued, URLs visited for each query, and allowed users to organize and annotate the saved information. Qvarfordt et al. [34] created a browser extension called *SearchPanel* that visualized users' document and process metadata (e.g., search engine, websites, result lists, visiting status, current-selected page, number of visits) to help users quickly navigate and re-find documents during exploratory searches. Donato et al. [15] at Yahoo! developed and launched the *SearchPad* as a web browser extension to assist users in keeping track of related searches for the same *research mission* (i.e., for a specific exploratory search topic) and to enable users to take notes (on the site) and to share with others.

The *SearchPad* tool also allowed users to retrieve/re-find queries and webpages associated with their notes.

**2.1.2 Tools to support note-taking and synthesis.** To conduct learning activities (e.g., comparison, analysis, evaluation, discovery, integration/synthesis) during an information seeking process, users need to actively identify important concepts and determine relationships among the acquired information [27]. However, studies found that users often encounter problems with these tasks, including difficulties navigating between different documents, keeping track of materials, taking notes, and maintaining awareness of retrieved information and their own thoughts through the process [38]. Active reading strategies such as highlighting, annotating content, and reflecting have been shown to have benefits when engaging in complex tasks [32]. For instance, these strategies can help users with metacognitive monitoring of their search/learning process [39].

Scaffolding (e.g., prompts, step-by-step guides) has been shown to be an effective method to assist the learning process. Scaffolding can provide users assistance with planning, monitoring, evaluating, and other types of metacognitive activities. For example, prompts shown to users might help them reflect on what information to consider or prioritize as being important [20]. Zhang and Quintana [43] designed *Digital IdeaKeeper* to help students engage more deeply with content when they conduct online inquiries related to learning science topics. The *Digital IdeaKeeper* outlines four activity spaces (i.e., planning, searching, analyzing, synthesizing) and highlights activities (i.e., skim, read, and summarize) with corresponding prompts. Results from a study [43] showed that the tool can encourage students to engage in more cognitive and metacognitive activities that might not happen otherwise: students searched more efficiently and effectively by planning and monitoring their search process, and they were better self-regulated for articulating their learning goals and keeping track of their task progress.

### 2.2 Metacognition

Exploratory search also includes metacognitive aspects (i.e., "thinking about thinking" [17]) as searchers engage in the search and learning processes. Metacognition refers to "the conscious self-awareness of one's own knowledge of task, topic, and thinking, and the conscious self-management (executive control) of the related cognitive processes." [30, p.364] (describing [22]).

In the education community, researchers have found that students often do not engage in metacognitive activities of planning, evaluation, and monitoring without encouragement or support [1, 2, 42]. To address this, researchers have developed and tested approaches to encourage students to engage in metacognitive activities as part of an active learning process. A variety of approaches have been shown to encourage metacognitive behaviors, including educational tools/software features, targeted instruction about metacognition, and the use of scaffolding to help guide learners [1, 35, 44]. These approaches have also shown learning and outcome benefits across a variety of task types and topic domains. For example, An and Cao [1] found that metacognitive scaffolding involving a planning sheet and question prompts improved students' planning skills and helped their design problem solving processes. Benefits of encouraging metacognitive activities has also been found

in scientific literacy [30], mathematical problem solving [8, 29], ill-structured problem solving [1], and topic understanding tasks.

Research has also indicated potential benefits of encouraging people engaged in search-based learning tasks to engage in metacognitive activities [2]. Stadler and Bromme developed a tool called *met.a.ware*, designed to encourage users to engage in metacognitive activities and information structuring while doing research and taking notes about an assigned medical topic [37]. Participants who received metacognitive evaluation prompts performed better on post-task assessments about knowledge of sources. They also found that providing ontological categories encouraged participants to save more information and to better structure their notes.

Huertas et al. [21] developed and evaluated metacognitive scaffolding for students working on an information search task. The scaffolding included prompts to create and modify a search plan, evaluate found information, and to monitor understanding and progress. Results showed that students who completed the tasks with the metacognitive scaffolding performed better on a post-search learning achievement evaluation [21].

Kuo et al. [25] developed and evaluated a tool called *Meta-Analyzer* that prompted students through a web search problem solving task by posing a series of "prompt" questions to guide the searcher. Results showed that participants who used the tool performed better as compared to a control group who used a conventional approach.

Crescenzi [11] investigated metacognitive activities described by participants while searching and found evidence of metacognitive regulation activities including orienting, monitoring, adaptation, and evaluation.

These previous efforts suggest that there are potential benefits of encouraging searchers to engage in metacognitive activities during learning-oriented searches. Our work explores a "lightweight" approach to encouraging metacognitive activities that does not require the system to have domain-specific knowledge.

### 3 SYSTEM DESIGN

For each task, participants interacted with two system components: (1) the search system, and (2) an information organization tool (the OrgBox or OrgDoc).

#### 3.1 Search System

To search for information, we provided participants with a custom-built search system (left of Figure 1). The system provided access to the LDC New York Times corpus of over 1.8 million news articles spanning the years 1987–2007. A traditional search interface powered by Lucene was provided that allowed participants to enter a query and then be presented with a ranked list of results. Results were presented in a standard search result page format, with 10 results per page. The interface included pagination controls at the bottom of the page to explore up to 50 results. Each result consisted of a title, URL, and a query-biased text snippet from the result landing page. Clicking a result would display a landing page with a text-only version of the NYT news article. Text from a landing page could be drag-dropped and/or copy-pasted from the landing page into the OrgTools. In the OrgBox tool, these actions automatically copied over the page title and URL in addition to the copied text.

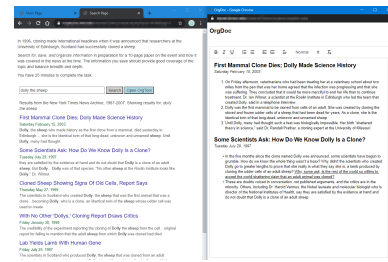


Figure 1: View of the search system (left) and the OrgDoc tool (right) as participants were instructed to arrange their browser windows.

The system logged all user interactions (e.g., queries, clicks, mouse events, scrolls). The search system presented the task description at the top of the page and included a button to launch the OrgTool associated with the current task (i.e., OrgBox or OrgDoc).

#### 3.2 OrgDoc Tool

The OrgDoc tool (right of Figure 1) was developed as a baseline system for saving and organizing information against which to compare the OrgBox. It provided familiar features common in rich-text editors and online document editing tools such as Google Docs. The tool provided a multi-line text box in which participants could enter and format text. Text could be directly typed into the window, or could be drag-dropped or copy-pasted from landing pages. The OrgDoc provided a number of common text formatting features including: bold, italic, underline, multi-level indentation, and bullet items. User interactions with the OrgDoc were logged (e.g., drag-drop events, copy-paste events, formatting events).

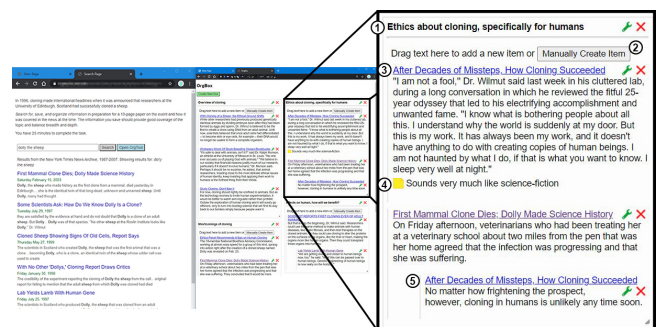


Figure 2: View of the search system (left) and OrgBox tool (center, right). The OrgBox displays (1) box title, (2) "Manually Create Item" button, (3) item title, URL (as link), edit and remove buttons, and text, (4) item note/annotation, and (5) nested items

#### 3.3 OrgBox Tool

The OrgBox tool (Figure 2) was developed to provide an easy-to-use set of information organization features. The feature set was based on results from a previous study on note-taking behaviors during search [12]. More detail about the OrgBox can be found in [40].

Participants could organize information *items* into groups called *boxes*. Boxes could be created either by clicking a "Create New Box" button at the top of the tool, or by drag-dropping text from a landing page into a blank area on the OrgBox canvas. When boxes were created, a dialog box displayed to allow users to enter a title

for the box. Box titles could be edited by clicking a wrench icon at the top of the box. Boxes could be deleted by clicking an “X” icon in the upper-right corner and then confirming the delete action.

Information *items* could be created by either drag-dropping them from a landing page into a box, or by clicking the “Manually Create Item” button at the top of each box. Items created through drag-drop automatically had the selected landing page text, page title, and a URL saved and displayed with the item. Clicking the URL displayed for the item would cause the main search window to navigate back to the landing page for the corresponding news article. Clicking the “Manually Create Item” button displayed a dialog box that allowed participants to enter text, a page title, URL, and an *item note* (an optional annotation) for the item. A wrench icon at the top of each item allowed participants to manually edit the item’s title, URL, text, and item note. Items could be re-arranged within and across boxes by drag-drop. Items could also be nested (indented) within other items with a maximum nesting level of 3. Items could be edited and deleted using the wrench and “X” controls at the top of each item.

## 4 METHOD

### 4.1 Study Overview & Protocol

To address our RQs, we conducted a remote user study with two within-subjects independent variables: OrgTool (OrgDoc vs. OrgBox) and task topic (Hurricane, Cloning). 24 participants were recruited from our university community and included undergraduate ( $n=11$ ) and graduate students ( $n=11$ ) as well as staff ( $n=2$ ).<sup>1</sup> Participants ranged in age from 18 to 35 years ( $M=23.8$ ,  $SD=5.07$ ). Participants self-reported their gender: female (58%,  $n=14$ ), male (42%,  $n=10$ ), or in a “self-identify” text field ( $n=0$ ).

The study was conducted remotely using Zoom video conferencing software and lasted about 2 hours. The study protocol proceeded as follows. The moderator introduced themselves, made sure Zoom was working, obtained informed consent, asked participants to share their screen in Zoom (after closing windows they did not wish to share), and asked participants to verbally agree to the session and their screen being recorded. Next, the moderator guided the participant through the study instructions and a practice task to have them practice “thinking out loud” and dragging-and-dropping text into a basic OrgTool (without boxes or a text editor).

Participants then completed two experimental tasks that followed the same sequence of steps outlined below. First, participants watched a short video introducing the search system and the OrgTool for the task. Next, participants completed a pre-task questionnaire which included the task description. The system then showed the search system and participants began their tasks; they could query, examine results, view articles, and interact with the OrgTools as they wished. Participants were instructed to think-aloud as they completed the tasks (think-aloud data is not analyzed in this paper). While the participants worked on the task, the moderator turned off their own video and microphone so as not to disturb the participant’s work.

<sup>1</sup>Power analysis was conducted during study design; however, no prior study results or effect sizes were available to inform assumptions before the study was run. Power analysis indicated a sample size of 24 was adequate based on the means of an average of all metacognitive items for pilot test participants. ( $OD_M=5.4$  and  $OB_M=6.5$ ) with a correlation of .5 between measures.

Table 1: Scenario, task topics, and task instructions displayed for each task.

---

**Scenario:** Imagine that you are taking a class that requires you to write 10-page papers on significant world events and how these events were covered in the news in the past.

---

**Task topics** (one per task):

**Hurricane:** In 2005, Hurricane Katrina struck the United States Gulf Coast causing massive destruction and great loss of life.

**Cloning:** In 1996, cloning made international headlines when it was announced that researchers at the University of Edinburgh, Scotland had successfully cloned a sheep.

---

**Task instructions** Search for, save, and organize information in preparation for a 10-page paper on the event and how it was covered in the news at the time. The information you save should provide good coverage of the topic and balance breadth and depth. You have 25 minutes to complete the task.

---

Participants were given up to 25 minutes to work on each task; they were told that they should let the moderator know when they were done. If the participant was still working after 20 minutes, the moderator gave them a visual 5-minute warning. After each search task, participants completed a post-task questionnaire.

After completing both tasks, participants completed an exit questionnaire and were offered a break. An exit interview about their experiences (not reported here) lasted about 30 minutes. Finally, participants were thanked and offered a \$40 electronic Amazon gift card.

### 4.2 Tasks

We used a simulated work task scenario [4] in the study instructions to ask participants to imagine that they were taking a class that required them to write papers about significant world events and how the events were covered by the news in the past. We asked participants to search for, save, and organize information into a “set of notes” to help them prepare to write a 10-page paper about two topics we gave them (Hurricane Katrina, the cloning of Dolly the Sheep), but we did not ask them to write a paper. Participants were told that they could access a news article search system and an information organization tool to use to save and organize information. To provide guidance, participants were asked to be creative and include subtopics they thought were interesting and important to cover while balancing depth and breadth. The scenario, topics, and task descriptions are shown in Table 1. Participants were randomly assigned to counterbalanced combinations of OrgTool and topic.

### 4.3 Data collection

Data collection involved logged system interactions, pre-task and post-task questionnaires, and an exit questionnaire at the end.

**4.3.1 Logged system interactions.** The search system logged queries, clicks, scroll events, and drag-drop events. Both OrgTools logged drag and drop events (from the search system and within the tool). The OrgBox also logged manual additions of boxes and items (i.e., using the buttons); changes in position of boxes and items; and the text of all additions, modifications, or deletions and its

corresponding element (e.g., title, URL, text, item note). In addition to drag and drop events, the OrgDoc system logged paste events (from the search system and within the tool), and the use of each formatting feature (i.e., bullet, number, bold, italics, underline, and change alignment).

**4.3.2 Questionnaires.** Unless otherwise noted, all questionnaire items asked participants to indicate their level of agreement with statements using a 7-point scale with labeled endpoints (1=strongly disagree, 7=strongly agree). The full text of questionnaire items is shown in Table 4; composite variables were created from multiple items as described in Section 4.4.1.

The pre-task questionnaire contained seven items covering aspects of participants' perceptions of their interest in the task, prior knowledge about the topic, and difficulty of the task (3 items).

The post-task questionnaire asked participants to indicate their level of agreement with 31 statements (see Table 4) about the *perceived cognitive and metacognitive benefits of using the OrgTool* (13 items), *task difficulty* (3 items), *workload* (6 NASA-TLX items) [19], the *quality and coverage* of the set of notes they created (4 items), the *OrgTool usability* (4 items), the *search system usability* (4 items), and the relevance of the search results (1 item). The *perceived cognitive and metacognitive benefits* items were adapted from [28] and included questions about the extent to which the tool helped them to: start and plan their search; organize and synthesize found information; track their progress; and evaluate their progress and search process. Two open-ended questions (not analyzed in this paper) asked participants to explain *why* they assigned each of the grades.

An exit questionnaire contained four open-ended questions asking participants to describe what they liked and disliked about each OrgTool. Participants were also asked demographic questions: age, University status (i.e., undergrad, grad student, staff, faculty), and gender (female, male, self-identify).

## 4.4 Data analysis

We conducted quantitative and qualitative analysis on participants' responses to items on the pre-task, post-task, and exit questionnaires. Stata 15 was used for quantitative analysis.

**4.4.1 Creating measures.** We created measures corresponding to (1) search and OrgTool interaction measures and wordcounts from logged interactions, (2) features of the final sets of notes, and (3) composite variables from the post-task questionnaire.

From the logged data, we created measures to enable comparison between OrgTool conditions:

- (1) time spent on task
- (2) *search behaviors*: # queries issued, # articles viewed, avg. # articles viewed per query, max rank of SERP mouse hover, max rank of articles viewed from SERP
- (3) *information saving*: # times text added to the OrgTool from articles (# drag or paste, # drag, # adds per article), # words in adds (total, average # words per add).
- (4) *information organization*: # words in final set of notes, # words per section, # article titles added, # notes with URLs.

We also calculated measures specific to each OrgTool. For OrgBox, these include the number of boxes and items created or moved within the set of notes. For the OrgDoc, this includes the number of

text adds into the document from an article (drag or paste), moves within notes (drag or paste), and uses of each formatting feature.

**Content analysis of notes.** The final set of notes created by the participants were coded for the information structures they contained using qualitative analysis methods. Two authors (A1, A3) created a coding guide of notes structures using an iterative, inductive coding approach with multiple rounds of open coding and discussion to arrive at a final coding guide. The final coding guide contained: (1) # *major sections*. For the OrgBox, we counted the number of boxes. For the OrgDoc, we counted the number of headings (i.e., text that was visibly distinct such as bold, italicized, or outdented), (2) # *article titles* included in notes, (3) # *of notes containing URLs*.

Two authors (A2, A3) coded the final set of OrgBox and OrgDoc notes. The sampling unit was the box or item in the OrgBox condition, and the line of text in the OrgDoc condition. For the first set of notes coded (OrgDoc), two researchers independently coded 4 sets of notes and had complete agreement (Krippendorff's  $\alpha=1$ ). The two researchers then each independently coded 10 sets of notes. For the second set of notes coded (OrgBox), one researcher (A3) independently coded all 24 sets of notes and another researcher (A2) reviewed all codes and noted disagreements. The reviewing process involved the researcher coding the notes and comparing the results. Disagreements were resolved by consensus.

**Composite questionnaire measures.** For constructs with items from existing instruments (with only minor adaptation), we assessed reliability using Cronbach's alpha and averaged item responses to create composite variables when  $\alpha > .65$ . For the 13 metacognitive support items, we created four factors based on the results of an exploratory factor analysis: *monitoring/tracking notes coverage* (3 items,  $\alpha=.73$ ), *metacognitive evaluation* (5 items,  $\alpha=.80$ ), *understanding the topic* (2 items,  $\alpha=.86$ ), and *metacognitive planning* (3 items,  $\alpha=.72$ ).<sup>2</sup>

**4.4.2 Testing for differences by OrgTool.** To test for differences in the OrgTool conditions, we used contrasts of the predicted values of the dependent variable estimated after multilevel mixed-effects models with a random intercept (i.e., a random effect for participant) and fixed effects of OrgBox, task sequence, and task topic. Restricted maximum likelihood was used with Kenward-Rogers degrees-of-freedom correction [24] for small sample inference. Multilevel mixed-effects Poisson or Negative Binomial models were estimated for count dependent variables (e.g., counts of search or saving behaviors). Test statistics for contrasts of marginal predictions are shown in Tables 2-4 with uncorrected *p*-values.

## 5 RESULTS

We present the results of our analysis by research question focusing on the differences between OrgBox and OrgDoc (means, significance of contrasts). See Tables 2-4 for full descriptive statistics and test statistics of the contrast of predicted margins of OrgBox vs. OrgDoc after multilevel models as described in Section 4.4.

<sup>2</sup>The exploratory factor analysis results are available from the authors and online: <https://doi.org/10.17615/c7b0-2t77>.



Table 2: Search behaviors (RQ1). Descriptive statistics by OrgTool conditions (mean, SD). Test statistic for contrasts of marginal linear predictions of OrgBox vs. OrgDoc conditions after multilevel mixed-effects negative binomial models ( $\chi^2(1)$ ). \*  $p < .05$ , †  $p < .01$ , ‡  $p < .001$ .

measure	OrgDoc <i>M (SD)</i>	OrgBox <i>M (SD)</i>	contrast $\chi^2(1)$
<b>RQ1. Search behaviors</b>			
# queries	4.08 (3.00)	5.46 (3.35)	2.77
# articles viewed	9.83 (3.36)	11.54 (6.87)	1.86
articles per query	<b>4.05</b> (3.67)	<b>2.50</b> (1.34)	* <b>4.44</b>
max hover rank	26.21 (14.74)	26.46 (14.98)	.02
max article rank	21.21 (13.79)	18.17 (11.21)	1.56
<i>N</i>	24	24	

## 5.1 Search behaviors (RQ1)

RQ1 focuses on understanding if there are differences in search behaviors between tasks completed with the OrgBox vs. the OrgDoc tools. To address RQ1, we analyzed search interaction measures shown in Table 2 and measured as described in Section 4.4.1. There was not a significant difference in the time spent on tasks in the OrgTool conditions ( $OB_M=25.21$ ,  $OD_M=24.69$ ,  $F(1,21)=.52$ ).

Participants issued more queries ( $OB_M=5.46$ ,  $OD_M=4.08$ ), and viewed more articles ( $OB_M=11.54$ ,  $OD_M=9.83$ ) for tasks completed using the OrgBox; although these differences were not statistically significant. However, there were significantly fewer articles viewed per query with the OrgBox ( $OB_M=2.50$ ,  $OD_M=4.05$ ).<sup>3</sup> There were no significant differences in the maximum rank of mouse hovers ( $OB_M=26.46$ ,  $OD_M=26.21$ ) or rank of article views from the SERP ( $OB_M=18.17$ ,  $OD_M=21.21$ ).

## 5.2 Information saving and organizing (RQ2)

RQ2 investigates differences in information saving and organizing behaviors for tasks completed with the OrgBox vs. the OrgDoc. We analyzed information saving and organizing measures derived from system logs and the features of final notes (see Table 3). We compared measures between systems when comparable (e.g., number of text add events) and described measures appropriate for only one system (e.g., OrgDoc formatting).

**5.2.1 Information saving.** We analyzed the number of times participants added to their set of notes directly from the SERP or an article. There was not a significant difference in the total number of add events between OrgTool systems ( $OB_M=20.21$ ,  $OD_M=19.79$ ); however, participants had significantly more adds by dragging text for OrgBox tasks ( $OB_M=20.21$ ) than OrgDoc tasks ( $OD_M=13.50$ ).<sup>4</sup> We did not find a significant difference in the average number of times text was added per article ( $OB_M=2.42$ ;  $OD_M=2.23$ ) or added per query ( $OB_M=5.30$ ;  $OD_M=7.75$ ).

We also looked at the number of words added directly from the SERP or a document. There were significantly more words added

<sup>3</sup>There were half as many participants for OrgBox tasks who issued two or fewer queries than for OrgDoc tasks ( $OB_n=4$ ,  $OD_n=8$ ) and they viewed fewer articles ( $OB_M=6.5$ ,  $OD_M=11$ ) and articles per query ( $OB_M=8.19$ ,  $OD_M=3.88$ ).

<sup>4</sup>Although participants were instructed to drag and drop text into the OrgTools, they also used keyboard shortcuts to copy and paste text for OrgDoc ( $OD_M=6.29$ ). This was likely due to participants' familiarity with using copy and paste to add text into other text editors that they have used in the past.

Table 3: Information saving and organizing measures (RQ2) derived from logs or codes. Descriptive statistics by OrgTool conditions (mean, SD). Test statistic for contrast of predicted counts (marginal linear predictions) of OrgBox vs. OrgDoc conditions after multilevel mixed-effects negative binomial models ( $\chi^2(1)$ ). \*  $p < .05$ , †  $p < .01$ , ‡  $p < .001$ .

measure	OrgDoc <i>M (SD)</i>	OrgBox <i>M (SD)</i>	contrast $\chi^2(1)$
<b>RQ2. Info saving to OrgTool from search system</b>			
# total text adds	19.79 (11.09)	20.21 (7.47)	.23
# text drags	<b>13.50 (10.76)</b>	<b>20.21 (7.47)</b>	‡ <b>16.72</b>
# text adds per article	2.23 (1.31)	2.42 (2.24)	.10
# adds per query	7.75 (6.33)	5.30 (4.15)	2.46
# total words	<b>981.38 (711.85)</b>	<b>1463.83 (804.48)</b>	‡ <b>14.94</b>
# words per add	<b>54.91 (47.75)</b>	<b>80.66 (52.76)</b>	‡ <b>12.98</b>
<b>RQ2. Organization in set of final notes</b>			
# sections	<b>4.04 (2.27)</b>	5.25 (1.62)	* <b>4.81</b>
# words in final	<b>966.71 (684.32)</b>	<b>1163.04 (672.19)</b>	* <b>5.35</b>
# words per section	218.30 (123.40)	235.40 (150.53)	.11
# article titles	<b>3.00 (4.36)</b>	<b>7.88 (5.59)</b>	‡ <b>7.14</b>
contained 1+ URL	n=2	n=24	
<i>N</i>	24	24	

to the OrgBox over the course of the task ( $OB_M=1463.83$ ) than the OrgDoc ( $OD_M=981.38$ ), and the average number of words included in an add event to the OrgBox was significantly greater than the OrgDoc ( $OB_M=80.66$ ,  $OD_M=54.91$ ).

**5.2.2 Information organizing.** For RQ2, we also looked to see if there were differences in how participants structured and organized their final set of notes between the OrgBox and OrgDoc systems.

As shown in Table 3, there was a significant difference in the number of sections in the final set of notes: notes in the OrgBox had an average of 5.25 sections (i.e., boxes) and OrgDoc notes had an average of 4.04 sections (i.e., top-level headings as differentiated by text formatting such as bold, italics, or indenting). Overall, there were significantly more words in the final set of notes for the OrgBox ( $OB_M=1163.04$ ) than OrgDoc ( $OD_M=966.71$ ) although there was not a significant difference in the number of words per section ( $OB_M=235.40$ ,  $OD_M=218.30$ ).

We also coded whether the sections were organized by articles or “themes” (e.g., negative effects, positive effects, event description). The majority of notes were organized by themes in both OrgTools ( $OB_n=23$ ,  $OD_n=21$ ) although there were some notes organized by articles ( $OB_n=0$ ,  $OD_n=2$ ) or articles and themes ( $OB_n=1$ ,  $OD_n=1$ ). Fisher’s exact test indicated there was not a significant difference in distribution between the two OrgTool conditions ( $p=.48$ ).

Finally, we looked the amount of information about the articles that were contained in the notes. In OrgDoc notes, 11 participants (46%) manually added article titles and two participants (8%) added full or partial URLs. For OrgBox tasks, article titles and URLs were automatically populated when a participant dragged text into the OrgBox. Unsurprisingly, we found significantly more article titles in the final set of notes for the OrgBox ( $OB_M=7.88$ ,  $OD_M=3.00$ ).

**5.2.3 OrgTool-specific behaviors.** For tasks completed with the OrgBox, the majority of boxes were created by participants clicking the “Create New Box” button ( $M=4.67$ ,  $SD=2.24$ ) versus dragging text into an empty space ( $M=1.04$ ,  $SD=1.73$ ). However, the

vast majority of items were created by dragging text into a box ( $M=19.17$ ,  $SD=7.35$ ) versus clicking the “Manually Create Item” button ( $M=1.29$ ,  $SD=2.61$ ). On average, participants added an item note when they created an item about three times per task ( $M=3.08$ ,  $SD=4.68$ ). In terms of rearranging the OrgBox contents, participants moved boxes within the canvas an average of 4.46 times ( $SD=4.54$ ) and items within boxes 7.67 times ( $SD=8.58$ ).

In OrgDoc tasks, participants added text to the OrgDoc from the SERP or an article an average of 19.79 times ( $SD=11.09$ ); they more frequently dragged text to the OrgDoc ( $M=13.50$ ,  $SD=10.76$ ) than copying and pasting ( $M=6.29$ ,  $SD=6.62$ ). Participants used text formatting features an average of 11.08 times during OrgDoc tasks ( $SD=11.35$ ); these included converting paragraph text to a bulleted or numbered list ( $M=4.38$ ,  $SD=4.69$ ), indenting ( $M=2.13$ ,  $SD=4.23$ ) or changing the appearance of the text by making bold ( $M=3.13$ ,  $SD=3.71$ ), using italics ( $M=.71$ ,  $SD=1.55$ ), or underlining ( $M=.63$ ,  $SD=1.24$ ).

### 5.3 Task, system, and notes perceptions (RQ3)

RQ3 investigates if there were differences between tasks completed with the OrgBox vs. the OrgDoc in how participants perceived (a) the usability of the search system and OrgTool, (b) the quality of the notes they produced, and (c) the topics and tasks. Table 4 summarizes these results.

**5.3.1 Usability of search and OrgTool systems.** To understand how the systems were perceived, we analyzed measures evaluating the OrgTools overall (i.e., grade on a 0-100 scale) and participants’ perceptions of the usability of the search system and each OrgTool. Participants gave significantly higher grades on a 0-100 scale to the OrgBox system ( $OB_M=85.58$ ) than to the OrgDoc system ( $OD_M=75.50$ ). Comparing the usability measures of tasks completed with the OrgDoc and OrgBox, we found no significant differences in perceived OrgTool usability ( $OB_M=5.57$ ;  $OD_M=5.52$ ), search system usability ( $OB_M=5.64$ ;  $OD_M=5.84$ ), or search results relevance ( $OB_M=4.83$ ;  $OD_M=4.96$ ).

**5.3.2 Quality of task product.** We did not find a difference in perceived notes quality between the OrgBox and OrgDoc tasks. There was not a difference in the overall assessment of notes quality (i.e., grade on a 0-100 scale) between the OrgTool systems ( $OB_M=78.25$ ,  $OD_M=77.63$ ). We also did not find a difference in perceived notes coverage: participants slightly agreed that they created very detailed and accurate notes ( $OB_M=4.08$ ,  $OD_M=4.25$ ) and covered at least one facet very deeply ( $OB_M=4.62$ ,  $OD_M=4.79$ ). They disagreed that they covered all facets of the topic ( $OB_M=2.92$ ,  $OD_M=3.12$ ) and covered all facets of the topic deeply ( $OB_M=2.42$ ,  $OD_M=2.46$ ).

**5.3.3 Topic and task perceptions.** We found no significant differences between the OrgBox and OrgDoc for any topic or task related perceptions from the pre- or post-task questionnaire (see Table 4 for full details). Overall, participants found the tasks to be interesting, reported low topic knowledge, and did not think they could create a set of detailed and accurate notes without searching. They did not expect that it would be difficult to complete the task or organize the information they found, and they neither agreed nor disagreed that it would be difficult to decide when to stop working on the task. After they completed the tasks, participants disagreed that

the task was difficult, it was difficult to search for information, and it was difficult to organize the information. The average workload as measured by the NASA-TLX was low ( $M=3.45$ ,  $SD=.82$ ).

### 5.4 Cognitive and metacognitive support (RQ4)

RQ4 focuses on whether there were differences in how the OrgBox and OrgDoc provided support for participants’ cognitive and metacognitive activities during the task using four measures of metacognitive support (see Section 4.4.1 for how these were created): monitoring and tracking notes coverage, evaluating process and task progress, understanding the topic, and planning. As shown in Table 4, when using the OrgBox (compared to the OrgDoc), participants reported significantly higher levels of support for monitoring/tracking notes coverage ( $OB_M=5.75$ ,  $OD_M=4.76$ ), metacognitive evaluation ( $OB_M=5.10$ ,  $OD_M=4.38$ ), and understanding the topic ( $OB_M=5.31$ ,  $OD_M=4.33$ ). Participants also rated the OrgBox higher in terms of metacognitive planning ( $OB_M=4.17$ ,  $OD_M=3.69$ ), however, this difference was not statistically significant ( $p=.059$ ).

## 6 DISCUSSION

We conducted a user study to investigate an information saving and organizing tool, the OrgBox, against a baseline rich-text editor called the OrgDoc. Results show that the OrgBox tool provided more cognitive and metacognitive support (RQ4) and was given a higher overall evaluation ratings (RQ3). The sets of notes created in the OrgBox contained more words and were divided into more sections (RQ2). Although the OrgBox tool was a novel tool, we did not find any difference between it and the OrgDoc in terms of perceived usability, workload, or task difficulty (RQ3). In addition, we did not find significant differences in search behaviors across the two tools (RQ1) except that users viewed slightly fewer articles per query when using the OrgBox tool. We discuss each of these results below in more detail.

**RQ1:** Regarding search behaviors (RQ1), we found that participants viewed significantly fewer articles per query in the OrgBox tasks although we did not find a significant difference in other search-related measures (i.e., number of queries, articles, max rank of SERP hovers or articles viewed, or the number of times text was added to the OrgTool per query). This suggests that, for the most part, participants did not interact with the search system differently using the OrgBox vs. the OrgDoc. We do however note that we observed a trend in our data that participants issued more queries and viewed more articles with the OrgBox than with the OrgDoc. While this trend did not reach statistical significance, it suggests an area for future work with a larger sample size.

**RQ2:** In terms of RQ2, participants created more “sections” of information in their notes and saved significantly more text overall and per text add from the search system when using the OrgBox compared to the OrgDoc. These findings suggest that the OrgBox encouraged people to create finer-grained knowledge structures about the topic during their search, and that they were encouraged to save more overall information into those structures. These results extend findings by Stadler and Bromme that found that providing tabbed ontological classification sections encouraged users to create more structured notes and to save more information compared to text-based note-taking [37]. In the Stadler and Bromme

Table 4: Participants' perceptions of OrgTool and search systems, notes quality, and task topic (RQ3) as well as cognitive and metacognitive benefits (RQ4) from pre- and post-task questionnaires. Descriptive statistics by OrgTool conditions (mean, SD), Cronbach's  $\alpha$  for composite variables; unless otherwise noted 1-7 scale where 1=strongly disagree, 7=strongly agree. Test statistic for contrasts of marginal linear predictions of OrgBox vs. OrgDoc conditions after multilevel mixed-effects model ( $F(1, 21)$ ) \*  $p < .05$ , †  $p < .01$ , ‡  $p < .001$ .

measure	OrgDoc <i>M (SD)</i>	OrgBox <i>M (SD)</i>	contrast <i>F(1, 21)</i>
<b>RQ3a: OrgTool and search system perceptions</b> (post-task)			
What grade would you give on a scale of 0-100 for the OrgTool system?	75.50 (16.01)	85.58 ( 9.49)	† 7.23
<i>Usability of OrgBox/OrgDoc</i> (composite, 4 items, $\alpha=.79$ )	5.52 (1.17)	5.57 ( .98)	.03
Thinking about your use of the OrgDoc/OrgBox: It worked the way I expected it to.			
...It was easy to use.			
...I experienced errors or had to redo steps because they didn't work the first time. (rev.)			
...It was a frustrating experience. (rev.)			
<i>Search system usability</i> (composite, 4 items as above asking about search system, $\alpha=.69$ )	5.84 ( .81)	5.64 (1.03)	1.67
It [search system] returned many results that were relevant to my task.	4.96 (1.43)	4.83 (1.61)	.11
<b>RQ3b: Notes quality and coverage</b> (post-task)			
What grade would you give on a scale of 0-100 for the notes you added to the OrgTool system?	77.63 (10.52)	78.25 (12.34)	.05
<i>Notes quality and coverage</i> (not combined, $\alpha=.647$ )			
I created a very detailed and accurate set of notes.	4.25 (1.29)	4.08 (1.32)	.35
I covered at least one facet very deeply.	4.79 (1.22)	4.62 (1.66)	.20
I covered all facets of this topic.	3.12 (1.39)	2.92 (1.50)	.35
I covered every facet of the topic very deeply	2.46 (1.22)	2.42 (1.32)	.02
<b>RQ3c: Topic and task perceptions</b> (pre- and post-task)			
I am interested to learn more about the topic of this task. (pre-task)	4.88 (1.57)	5.08 (1.10)	.28
I know a lot about this topic. (pre-task)	2.92 (1.74)	2.67 (1.34)	.35
I can create a very detailed and accurate set of notes now without needing to look for information.(pre-task)	1.67 (1.13)	1.50 ( .72)	.48
<i>Expected task difficulty</i> (pre-task) (not combined, $\alpha=.57$ ).			
I think it will be difficult to complete this task.	3.13 (1.26)	3.00 (1.35)	.12
I think it will be difficult to decide when to stop working on this task.	4.21 (1.72)	3.96 (1.68)	.99
I think it will be difficult to organize the information I find into the set of notes to complete this task.	3.17 (1.27)	3.00 (1.44)	.30
<i>Experienced task difficulty</i> . (post-task) (not combined, $\alpha=.53$ )			
It was difficult to complete this task.	3.12 (1.45)	3.00 (1.25)	.10
It was difficult to search for information to complete this task.	3.29 (1.43)	3.38 (1.53)	.04
It was difficult to organize the information I found into the set of notes to complete this task.	3.00 (1.69)	2.88 (1.23)	.08
<i>Workload</i> (post-task) (composite of 6 NASA-TLX items, $\alpha=.66$ )	3.53 ( .92)	3.37 ( .70)	1.38
<b>RQ4: OrgTool benefits</b> . (post-task)			
<i>Support for monitoring/tracking notes coverage</i> (composite of 3 questions, $\alpha=.73$ )	4.76 ( .97)	5.75 ( .84)	†13.86
The OrgDoc/OrgBox helped me to keep track of how broadly I was covering the topic.			
...keep track of how deeply I was covering some aspects of the topic.			
...organize the information about the topic.			
<i>Support for metacognitive evaluation of process, task progress</i> (composite of 5 questions, $\alpha=.80$ )	4.38 (1.12)	5.10 ( .93)	†10.32
...keep track of my progress towards completing the task.			
...decide when I had met the task goal.			
...evaluate whether my strategy for the task was working.			
...change my approach to the task.			
...ask myself if I had included important facets of the topic			
<i>Support for understanding the topic</i> (2 questions, $\alpha=.86$ )	4.33 (1.20)	5.31 (1.37)	* 6.85
...figure out how the information about the topic fits together.			
...determine facets of the topic.			
<i>Support for metacognitive planning</i> (composite of 3 questions, $\alpha=.72$ )	3.69 (1.03)	4.17 (1.38)	4.00 ( $p=.059$ )
...understand what the task was asking me to do.			
...develop a plan for approaching the task.			
...decide where to begin my search.			
<i>Average of all OrgTool benefit items</i> (composite of 13 items, $\alpha=.86$ )	4.30 ( .78)	5.07 ( .81)	† 12.79
<i>n</i>	24	24	



work, they provided users with pre-defined categories that were relevant to the search topic. In our study, we provided mechanisms to encourage structure creation, but did not provide pre-defined categories. However, we found similar results – users saved more information and created more structure (sections). This suggests that while it is helpful, systems need not provide pre-determined categories to users in order to encourage structuring. Rather, systems can encourage users to structure and save more information simply by providing mechanisms to do so that are well-integrated with the search interface.

Stadtler and Bromme reported that participants who used their text-based tool created considerably less structure on their own, and instead relied on the structure of the copied information [37]. In contrast, in our OrgDoc (text-based) condition, we observed that a majority of our participants did create their own structure for their saved information, but they created less structure (fewer sections) than in the OrgBox condition. We found that slightly less than half of the sets of OrgDoc notes contained article titles ( $n=11$ , 46%) and few ( $n=2$ , 8%) contained full or partial URLs. This was somewhat surprising as a study of hand-written notes [12] found article titles and full or partial URLs in 66% and 33% of the set of hand-written notes for oneself.

**RQ3:** For RQ3, we found that participants gave a higher overall evaluation rating to the OrgBox than the OrgDoc. We did not find significant differences in participants' perceptions of the tool usability, search system usability and quality, or task difficulty across the conditions. We interpret the lack of significant differences in usability and task difficulty as a positive. Even though the OrgBox tool was novel to the participants, they did not feel it was less usable, or that it made the task more difficult. These findings also suggest that the significant difference found in the overall rating (participants rated the OrgBox as higher overall) was not due to tool or search system usability, but was more likely due to the features of the OrgBox tool.

We also did not observe any difference based on OrgTool in participants' ratings of the quality or topic coverage of the notes they produced. Our RQ3 findings are consistent with results of a previous study which compared the OrgBox to a different baseline (i.e., a bookmarking tool) [9]. The results of that study also did not find differences in task difficulty, topic knowledge, or quality of the task output. Our participants gave their notes a mean overall quality assessment of 75 on a 0-100 scale, suggesting that they were generally satisfied with their notes. Although we had thought that the OrgBox might lead participants to produce notes that they viewed as "higher-quality," participants often self-rate their own task outputs as being generally satisfactory. It also may be the case that the OrgBox did not have a direct impact on the participants' perceptions of the quality of the notes even though it provided metacognitive support. For example, a study of metacognitive scaffolding to support ill-structured problem solving reported benefits on problem solving processes, but did not find differences in design outcomes [1].

**RQ4:** Perhaps our most interesting results are for RQ4. Participants reported that the OrgBox provided greater support for three specific types of cognitive and metacognitive activities: *understanding and synthesizing* the topic, *monitoring and tracking* the coverage

of the notes, and *evaluating* the task process and progress. Support for *metacognitive planning* almost reached significance ( $p=.059$ ).

These findings are important for several reasons. First, they are consistent with results of a previous study [9] which compared the OrgBox tool to a bookmarking tool baseline. The OrgDoc tool that we used as a baseline in the current study included features found in rich-text editors (e.g., Google Docs) that users commonly use to take notes while searching, whereas the bookmarking tool in [9] only supported saving information as bookmarks in linear list. Thus, the results of the current study provide significant evidence to show that the OrgBox tool provides metacognitive support to searchers over and beyond commonly used note-taking tools. Second, our results show that metacognitive support during search can be provided through lightweight, simple tools to help users save, organize, and re-structure information into groups. Prior work in both education and information seeking (e.g., [25, 37]) has explored tools that provide topic-specific scaffolding to help encourage metacognitive activities during learning. While these are important approaches, they require a search system to be able to provide topic and task-specific scaffolding. Our results suggest that topic-agnostic tools such as the OrgBox can also provide metacognitive support for searchers.

## 7 CONCLUSION

We conducted a user study to investigate the effects of an information saving and organizing tool, the OrgBox, against a baseline rich-text editor called the OrgDoc. Results show that participants reported that the OrgBox provided more cognitive and metacognitive support (RQ4) compared to the baseline tool. Participants saved more information (words) into their notes, and created more structures (sections) in their notes (RQ2) when using the OrgBox. Further, the OrgBox received higher overall evaluation ratings (RQ3) but did not increase the perceived difficulty of the tasks.

Our results provide evidence that tools that integrate information organization and synthesis with search can provide metacognitive support for users attempting to learn about a complex topic. Our results also suggest that such tool need not have domain-specific knowledge, nor be complex. Our novel OrgBox tool was easy for users to use to help them organize information as they wished. It encouraged users to save more information and to create more structures in their notes as compared against the baseline OrgDoc rich-text editor. This suggests that systems that integrate information seeking with tools to organize and save information into groups can be an effective way to encourage users to engage in more metacognitive activities during search. In addition, such tools could provide additional context for search algorithms to leverage and additional methods for search systems to interact with users.

## ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1552587. We thank our study participants for their time and the UNC Institutional Review Board.

## REFERENCES

- [1] Yun-Jo An and Li Cao. 2014. Examining the effects of metacognitive scaffolding on students' design problem solving and metacognitive skills in an online

- environment. *Journal of Online Learning and Teaching* 10, 4 (2014), 552–568. [https://jolt.merlot.org/vol10no4/An\\_1214.pdf](https://jolt.merlot.org/vol10no4/An_1214.pdf)
- [2] Roger Azevedo, Jennifer G Cromley, and Diane Seibert. 2004. Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology* 29, 3 (2004), 344–370. <https://doi.org/10.1016/j.cedpsych.2003.09.002>
  - [3] Marcia J Bates. 1979. Idea tactics. *Journal of the American Society for Information Science* 30, 5 (1979), 280–289. <https://doi.org/10.1002/asi.4630300507>
  - [4] Pia Borlund. 2003. The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research* 8, 3 (2003), 1–34. <http://informationr.net/ir/8-3/paper152.html>
  - [5] Leanne Bowler. 2010. A taxonomy of adolescent metacognitive knowledge during the information search process. *Library & Information Science Research* 32, 1 (2010), 27–42. <https://doi.org/10.1016/j.lisr.2009.09.005>
  - [6] Leanne Bowler. 2010. Talk as a metacognitive strategy during the information search process of adolescents. *Information Research* 15, 4 (2010), 1–13. <http://www.informationr.net/ir/15-4/paper449.html>
  - [7] Saskia Brand-Gruwel, Iwan Wopereis, and Amber Walraven. 2009. A descriptive model of information problem solving while using internet. *Computers & Education* 53, 4 (2009), 1207–1217. <https://doi.org/10.1016/j.compedu.2009.06.004>
  - [8] Maria Cardelle-Elawar. 1995. Effects of metacognitive instruction on low achievers in mathematics problems. *Teaching and Teacher Education* 11, 1 (1995), 81–95. [https://doi.org/10.1016/0742-051X\(94\)00019-3](https://doi.org/10.1016/0742-051X(94)00019-3)
  - [9] Bogeum Choi, Jaime Arguello, Robert Capra, and Austin R. Ward. 2021. OrgBox: A knowledge representation tool to support complex search tasks. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval (CHIIR '21)*. Association for Computing Machinery, New York, NY, USA, 219–228. <https://doi.org/10.1145/3406522.3446029>
  - [10] Keyvyn Collins-Thompson, Preben Hansen, and Claudia Hauff. 2017. Search as Learning (Dagstuhl Seminar 17092). *Dagstuhl Reports* 7, 2 (2017), 135–162. <https://doi.org/10.4230/DagRep.7.2.135>
  - [11] Anita Crescenzi. 2016. Metacognitive knowledge and metacognitive regulation in time-constrained information search. In *Search as Learning workshop at SIGIR '16*, Vol. 1647. 1–5. [http://ceur-ws.org/Vol-1647/SAL2016\\_paper\\_5.pdf](http://ceur-ws.org/Vol-1647/SAL2016_paper_5.pdf)
  - [12] Anita Crescenzi, Yuan Li, Yinglong Zhang, and Rob Capra. 2019. Towards better support for exploratory search through an investigation of notes-to-self and notes-to-share. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19)*. Association for Computing Machinery, New York, NY, USA, 1093–1096. <https://doi.org/10.1145/3331184.3331309>
  - [13] J. Shane Culpepper, Fernando Diaz, and Mark D. Smucker. 2018. Research frontiers in information retrieval: Report from the Third Strategic Workshop on Information Retrieval in Lorne (SWIRL 2018). *SIGIR Forum* 52, 1 (Aug. 2018), 34–90. <https://doi.org/10.1145/3274784.3274788>
  - [14] Brenda Dervin. 1998. Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of Knowledge Management* 2, 2 (1998), 36–46.
  - [15] Debora Donato, Francesco Bonchi, Tom Chi, and Yoelle Maarek. 2010. Do you want to take notes? Identifying research missions in Yahoo! Search Pad. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. Association for Computing Machinery, New York, NY, USA, 321–330. <https://doi.org/10.1145/1772690.1772724>
  - [16] Carsten Eickhoff, Jacek Gwizdka, Claudia Hauff, and Jiyin He. 2017. Introduction to the special issue on search as learning. *Information Retrieval Journal* 20, 5 (2017), 399–402. <https://doi.org/10.1007/s10791-017-9315-9>
  - [17] John H. Flavell. 1979. Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist* 34, 10 (1979), 906–911. <https://doi.org/10.1037/0003-066X.34.10.906>
  - [18] Jacek Gwizdka, Preben Hansen, Claudia Hauff, Jiyin He, and Noriko Kando. 2016. Search as Learning (SAL) workshop 2016. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. Association for Computing Machinery, New York, NY, USA, 1249–1250. <https://doi.org/10.1145/2911451.2917766>
  - [19] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human mental workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. Elsevier Science Publishers B. V., North Holland, 138–183.
  - [20] Janette. R Hill and Michael J Hannafin. 2001. Teaching and learning in digital environments: The resurgence of resource-based learning. *Educational Technology Research and Development* 49, 3 (2001), 37–52. <https://doi.org/10.1007/BF02504914>
  - [21] Adriana Huertas, Omar López, and Luis Sanabria. 2017. Influence of a metacognitive scaffolding for information search in B-Learning courses on learning achievement and its relationship with cognitive and learning style. *Journal of Educational Computing Research* 55, 2 (2017), 147–171. <https://doi.org/10.1177/0735633116656634>
  - [22] Janis E Jacobs and Scott G Paris. 1987. Children's metacognition about reading: Issues in definition, measurement, and instruction. *Educational Psychologist* 22, 3-4 (1987), 255–278. <https://doi.org/10.1080/00461520.1987.9653052>
  - [23] Diane Kelly, Jaime Arguello, and Robert Capra. 2013. NSF Workshop on Task-Based Information Search Systems. *SIGIR Forum* 47, 2 (Jan. 2013), 116–127. <https://doi.org/10.1145/2568388.2568407>
  - [24] Michael G. Kenward and James H. Roger. 1997. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53, 3 (1997), 983–997. <https://doi.org/10.2307/2533558>
  - [25] Fan-Ray Kuo, Nian-Shing Chen, and Gwo-Jen Hwang. 2014. A creative thinking approach to enhancing the web-based problem solving performance of university students. *Computers & Education* 72 (2014), 220–230. <https://doi.org/10.1016/j.compedu.2013.11.005>
  - [26] Gary Marchionini. 1995. *Information Seeking in Electronic Environments*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511626388>
  - [27] Gary Marchionini. 2006. Exploratory search: From finding to understanding. *Communications of the ACM* 49, 4 (2006), 41–46. <https://doi.org/10.1145/1121949.1121979>
  - [28] Lindsay McCardle and Allyson F. Hadwin. 2015. Using multiple, contextualized data sources to assess learners' perceptions of their self-regulated learning. *Metacognition and Learning* 10, 1 (2015), 43–75. <https://doi.org/10.1007/s11409-014-9132-0>
  - [29] Zemira R Mevarech. 1999. Effects of metacognitive training embedded in cooperative settings on mathematical problem solving. *The Journal of Educational Research* 92, 4 (1999), 195–205. <https://doi.org/10.1080/00220679909597597>
  - [30] Tova Michalsky, Zemira R Mevarech, and Liora Haibi. 2009. Elementary school children reading scientific texts: Effects of metacognitive instruction. *The Journal of Educational Research* 102, 5 (2009), 363–376. <https://doi.org/10.3200/JOER.102.5.363-376>
  - [31] Dan Morris, Meredith Ringel Morris, and Gina Venolia. 2008. SearchBar: A search-centric web history for task resumption and information re-finding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. Association for Computing Machinery, New York, NY, USA, 1207–1216. <https://doi.org/10.1145/1357054.1357242>
  - [32] Donatella Persico and Karl Steffens. 2017. Self-regulated learning in technology enhanced learning environments. In *Technology Enhanced Learning: Research Themes*. Springer International Publishing, 115–126. [https://doi.org/10.1007/978-3-319-02600-8\\_11](https://doi.org/10.1007/978-3-319-02600-8_11)
  - [33] Peter Pirolli and Daniel M. Russell. 2011. Introduction to this Special Issue on Sensemaking. *Human-Computer Interaction* 26, 1-2 (2011), 1–8. <https://doi.org/10.1080/07370024.2011.556557>
  - [34] Pernilla Qvarfordt, Simon Tretter, Gene Golovchinsky, and Tony Dunnigan. 2014. SearchPanel: Framing complex search needs. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '14)*. Association for Computing Machinery, New York, NY, USA, 495–504. <https://doi.org/10.1145/2600428.2609620>
  - [35] Ido Roll, Natasha G Holmes, James Day, and Doug Bonn. 2012. Evaluating metacognitive scaffolding in Guided Invention Activities. *Instructional Science* 40, 4 (2012), 691–710. <https://doi.org/10.1007/s11251-012-9208-7>
  - [36] Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. 1993. The cost structure of sensemaking. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. Association for Computing Machinery, New York, NY, USA, 269–276. <https://doi.org/10.1145/169059.169209>
  - [37] Marc Stadler and Rainer Bromme. 2008. Effects of the metacognitive computer-tool met. a. ware on the web search of laypersons. *Computers in Human Behavior* 24, 3 (2008), 716–737. <https://doi.org/10.1016/j.chb.2007.01.023>
  - [38] Craig S. Tashman and W. Keith Edwards. 2011. Active reading and its discontents: The situations, problems and ideas of Readers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2927–2936. <https://doi.org/10.1145/1978942.1979376>
  - [39] Shang Wang, Deniz Sonmez Unal, and Erin Walker. 2019. MindDot: Supporting effective cognitive behaviors in concept map-based learning environments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3290605.3300258>
  - [40] Austin Ward and Rob Capra. 2021. OrgBox: Supporting cognitive and metacognitive activities during exploratory search. In *The 44th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*. Association for Computing Machinery, New York, NY, USA, 4. <https://doi.org/10.1145/3404835.3462790>
  - [41] Ryen W White and Resa A Roth. 2009. Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 1, 1 (2009), 1–98. <https://doi.org/10.2200/S00174ED1V01Y200901ICR003>
  - [42] GE Xun and Susan M Land. 2004. A conceptual framework for scaffolding III-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research and Development* 52, 2 (2004), 5–22. <https://doi.org/10.1007/BF02504836>
  - [43] Meilan Zhang and Chris Quintana. 2012. Scaffolding strategies for supporting middle school students' online inquiry processes. *Computers & Education* 58, 1

- (2012), 181–196. <https://doi.org/10.1016/j.compedu.2011.07.016>
- [44] Michal Zion, Tova Michalsky, and Zemira R Mevarech. 2005. The effects of metacognitive instruction embedded within an asynchronous learning network on scientific inquiry skills. *International Journal of Science Education* 27, 8 (2005), 957–983. <https://doi.org/10.1080/09500690500068626>