

# The Effects of Working Memory during Search Tasks of Varying Complexity

Bogeum Choi, Robert Capra, Jaime Arguello  
University of North Carolina at Chapel Hill  
{choiboge,rcapra,jarguello}@unc.edu

## ABSTRACT

We report on a study that evaluated the effects of working memory and task complexity on participants' perceptions, behaviors, and outcomes. Twenty-four participants performed two search tasks of varying complexity and completed a psychometric test to measure working memory ability. Our results found several important trends. First, task complexity had an effect on participants' perceptions about temporal demand and satisfaction with the time spent on the task. Second, participants with higher working memory exerted more search effort (e.g., issued more queries). Third, participants with higher working memory had better outcomes, particularly during more complex tasks. Finally, while participants with lower working memory exerted less effort (engaged in satisficing behaviors) and had weaker outcomes, working memory did not affect participants' post-task perceptions about workload and satisfaction. We discuss implications of our results for developing search tools to support users with varying levels of working memory.

## KEYWORDS

Working memory; task complexity; search behavior

### ACM Reference Format:

Bogeum Choi, Robert Capra, Jaime Arguello. 2019. The Effects of Working Memory during Search Tasks of Varying Complexity. In *2019 Conference on Human Information Interaction and Retrieval (CHIIR '19)*, March 10–14, 2019, Glasgow, Scotland Uk. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3295750.3298948>

## 1 INTRODUCTION

In recent years, there has been a growing interest in better understanding how characteristics of individual users can influence search behaviors and outcomes. Studies have considered personality traits such as need for cognition (i.e., a person's disposition to engage with cognitively demanding activities) [32], as well as cognitive abilities such as perceptual speed [8, 28], visualization ability [8], visual memory [28], and working memory [16–18]. The end goal of this body of research is to better understand how search systems can best support users with different cognitive profiles.

We report on a laboratory study ( $N = 24$ ) that investigated the effects of working memory and task complexity on participants' perceptions, behaviors, and outcomes. Participants in the study

completed two *comparative tasks* of varying levels of complexity. Each task asked the participant to compare a set of items (or alternatives) along a set of dimensions (or attributes). To complete each task, participants were instructed to: (1) use a custom-built search engine to find information for all items and dimensions given in the task description, (2) enter information into a table consisting of items (rows) and dimensions (columns), (3) make a final item selection, and (4) provide a justification/explanation for their choice. Our manipulation of task complexity involved varying the number of items and dimensions participants were explicitly asked to consider as part of the comparative task (i.e.,  $2 \times 2$  vs.  $4 \times 4$ ).

Our study investigated the following three research questions:

**RQ1:** What are the effects of working memory and task complexity on participants' post-task perceptions about their performance? Specifically, we focused on participants' perceived levels of workload and satisfaction with their task outcomes and strategies.

**RQ2:** What are the effects of working memory and task complexity on participants' search behaviors? Specifically, we focused on measures associated with the amount of search effort and the pace of interaction.

**RQ3:** What are the effects of working memory and task complexity on participants' search outcomes? Specifically, we focused on participants' ability to address all combinations of items/dimensions associated with the task and the justification length.

## 2 RELATED WORK

Our research builds on prior studies on the effects of working memory and task complexity on search behaviors and outcomes.

**Working Memory:** Working memory is defined as someone's ability to hold information in short-term memory and to work with information that is no longer perceptually present [13]. Working memory is said to be a core executive function that is critical for making sense of anything that unfolds over time (e.g., inferring relations between items and ideas) [13]. In this respect, working memory impacts higher-level cognitive skills such as reasoning [5, 23], reading comprehension [12, 24] and problem-solving [19, 31].

Several studies have considered the impact of working memory on search behaviors. Gwizdka [16, 17] investigated the effects of working memory (WM) and task complexity on participants' behaviors while interacting with two different search interfaces. Both interfaces returned a ranked list of search results, but one interface (the *overview* interface) also presented tag clouds with clickable terms for users to filter the search results. In less demanding situations (i.e., during simple tasks using the *overview* system), high-WM participants completed tasks faster than low-WM participants. Conversely, in more demanding situations (i.e., during complex tasks using the baseline system), high-WM participants expended more effort than low-WM participants. In a later study, Gwizdka [18] compared the search behaviors of high- and low-WM

© {Owner/Authors} 2019. This is the authors' version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published as shown below.

participants while performing complex search tasks. While high- and low-WM participants had similar completion times, high-WM participants performed more actions and spent more time reading content pages (especially towards the end of the search session). Consistent with previous findings [16, 17], these results suggest that low-WM participants may be more likely to engage in satisficing behaviors during mentally demanding tasks.

Prior studies have also compared the search behaviors of dyslexic and non-dyslexic searchers [14, 25]. While dyslexia is a heterogeneous condition, evidence suggests a strong correlation between dyslexia and lower phonological working memory (i.e., the ability to hold words in short-term memory) [6, 7, 11, 25, 26]. MacFarlane *et al.* [25] examined the search behaviors of dyslexic and non-dyslexic students, and found that non-dyslexic students examined more documents and judged a greater percentage of documents as non-relevant. In other words, non-dyslexic students were more “selective” in their relevance judgments. Additionally, Fourney *et al.* [14] compared relevance judgments from dyslexic versus non-dyslexic study participants, and found that relevance judgments from dyslexic participants had a greater central tendency (i.e., were less bi-modal). Consistent with MacFarlane *et al.* [25], non-dyslexic participants were more “selective” in their relevance judgments.

**Task Complexity:** Task complexity has been defined and manipulated from different perspectives [29]. Campbell [9] characterized task complexity in terms of: (1) the number of required outcomes, (2) the number possible paths to the outcomes, (3) the level of uncertainty about the paths, and (4) the degree of interdependence between the paths. Jansen [21] proposed that search task complexity can be viewed through the lens of *cognitive complexity*, which relates to the amount of mental effort and learning required to complete the task. Jansen *et al.* [21] and Kelly *et al.* [22] used Anderson and Krathwohl’s taxonomy of learning outcomes [2] to develop tasks of varying complexity levels. Using this framework, studies have found that more cognitively complex tasks have greater levels of expected pre-task difficulty [10, 20, 22], experienced post-task difficulty [3, 10, 20, 22], search effort [3, 10, 20–22], and greater diversity of query reformulation strategies [30].

In this paper, we used comparative tasks, which fall under *analyze* tasks (intermediate complexity) within the cognitive complexity framework. Our manipulation of task complexity involved increasing the number of items and dimensions for participants to consider ( $2 \times 2$  vs.  $4 \times 4$ ). Based on Campbell’s view of task complexity [9], our  $4 \times 4$  tasks had more outcomes, more paths to the outcomes, and possibly more interdependence between paths.

### 3 USER STUDY

To investigate our three research questions, we conducted a laboratory study with 24 participants (18 females). Participants were recruited using an opt-in mailing list of employees at our university. Participants were asked to complete two comparative tasks that involved using a web search engine to find information. A comparative task is one that involves comparing a set of items across a set of dimensions. For example, a comparative “car shopping” task might require comparing different models across dimensions such as price, safety features, and fuel efficiency. During each comparative task, participants were given specific items and dimensions to consider. To complete each task, participants were required to

complete two sub-tasks. First, participants were asked to search for information in order to complete a table of  $n$  rows (items) and  $n$  columns (dimensions). Second, participants were asked to make a final item selection and write a justification for their choice based on the information found during their search.

**Independent Variables:** The study involved two independent variables: working memory and task complexity. Working memory was a between-subjects factor. As described below, we measured working memory capacity using the “operation-word-span” (OSPAN) test [27]. To study the effects of working memory (WM), participants were split into low- and high-WM groups using a median split. Task complexity was a within-subjects variable. We manipulated task complexity by varying the number of items and dimensions for participants to consider as part of the task. Simple tasks required comparing 2 items along 2 dimensions (i.e., a  $2 \times 2$  table), and complex tasks required comparing 4 items along 4 dimensions (i.e., a  $4 \times 4$  table). Each participant completed one simple and one complex task.

**Protocol:** First, the study moderator provided a general description of the study. Then, participants signed a consent form and completed a brief demographics questionnaire. To complete the tasks, participants were given access to a custom-built search engine to find information and a Microsoft Word document to enter their response. These two interfaces were displayed side-by-side. The custom-build search engine used the Bing Web Search API to retrieve results from the open web. The Word document provided a set of instructions and an empty table for participants to enter their response (i.e., a  $2 \times 2$  or  $4 \times 4$  table, depending on the task complexity condition). Header cells (top row and first column) in the given table contained the specific items and dimensions associated with the task. The custom-built search engine logged all user interactions on the search engine results page (SERP).

During each task, participants were instructed to find as much information as they thought was necessary to make an informed item selection. Participants were also asked to produce a written justification upon making a final decision. For each task, they were given 20 minutes with a 2-minute warning. Participants completed a post-task questionnaire after each task. After completing both tasks, participants were administered a working memory test that lasted 10-15 minutes. Participants were compensated \$15 USD for participating in the study. The study took about 60 minutes.

**Working Memory Test:** To measure working memory capacity, we used the computerized “operation-word-span” (OSPAN) test [27] distributed with the “CogLab in a CD” textbook by Francis *et al.* [15]. The OSPAN test measures an individual’s ability to recall words displayed in sequence while concurrently completing simple secondary tasks. Participants complete 15 trials of varying lengths. During each trial, the participant is shown a sequence of words and is then asked to recall the words in their original order from a grid display. Additionally, during each trial, participants complete simple math problems between each word shown in sequence (e.g., “Is  $(4/2)-1=1$ ?”). The final score is equal to the sum of sequence lengths of all trials perfectly recalled.

**Task Complexity Manipulation:** Our manipulation of task complexity involved varying the number of items and dimensions associated with the task (simple =  $2 \times 2$ , complex =  $4 \times 4$ ). To study the effect of task complexity, we developed four search tasks that varied

across two levels of complexity (simple vs. complex) and two topics (dogs breeds to consider for adoption vs. methods to quit smoking). Each participant was exposed to both task complexity levels and both task topics. The presentation order of task complexity and task topic was counter-balanced (i.e., the 4 possible orderings were repeated 6 times across our 24 participants). All task descriptions involved a scenario in which participants were asked to look for information for a friend. For example, our simple ( $2 \times 2$ ) “dog breed” task asked: “A friend of yours has recently decided to get a dog for companionship. Your friend works during the day and lives in an apartment complex. Now you are trying to help your friend make a decision. How do: (a) Pug and (b) Bichon Frise dog breeds differ as a choice for your friend in terms of (a) the ability to be left alone during the day, and (b) the need for outdoor activity?”

We chose to use comparative tasks for several reasons. First, comparative tasks are likely to involve working memory. A searcher must identify different items, search for information about specific items and dimensions, and compare information across the items and dimensions. This process requires processing new information while retaining prior information. Second, working memory is often considered in terms of the number of discrete units of information over which one can distribute attention and maintain an active state [4]. Therefore, we thought that increasing the number of elements to consider could increase memory load. Finally, the ability to structure comparative tasks as a table allowed us to measure search outcomes in a straightforward way (e.g., by considering the percentage of completed cells).

**Post-task Questionnaire (RQ1):** After each task, participants completed a 9-item questionnaire that measured workload and satisfaction with their performance. To measure workload, we included four items from the NASA-TLX questionnaire [1]: (1) mental demand, (2) temporal demand, (3) performance, and (4) effort. Additionally, we included five questions about participants’ satisfaction with their performance: (5) whether they found enough information, (6) satisfaction with the time spent on the task, (7) satisfaction with the amount of information found, (8) satisfaction with the quality of information found, and (9) satisfaction with the chosen search strategy. Participants responded to all items using a 5-point scale. For all workload items except for “performance”, high response values indicate high levels of workload. For all satisfaction items, high values indicate high levels of satisfaction.

**Behavioral Measures (RQ2):** To investigate the effects on participants’ search behaviors, we computed four behavioral measures: (1) number of queries issued, (2) number of SERP clicks, (3) average time to first SERP click after a query submission (in seconds), and (4) average time (in seconds) between subsequent search actions (i.e., queries and clicks).

**Outcome Measures (RQ3):** As previously mentioned, to complete the task, participants were required to fill a  $n \times n$  table, make an item recommendation, and provide a written justification. These notes were considered as an external representation of participants’ search performance other than behavioral measures. To investigate the effects on participants’ search outcomes, we considered two measures. First, our *coverage* measure considered the percentage of the  $n \times n$  table that was *not* left blank. Second, we considered the final justification length (in words) as a rough indicator of the justification’s quality (i.e., depth and/or breadth).

## 4 RESULTS

**Effects on Post-Task Perceptions (RQ1):** In RQ1, we investigate the effects of working memory and task complexity on participants’ post-task perceptions about workload and satisfaction. As shown in Table 1, mixed-effects ANOVAs found a significant main effect of task complexity on two measures: (1) temporal demand and (2) satisfaction with the time spent on the task. For complex tasks, participants reported greater levels of temporal demand and lower levels of satisfaction with the time spent on the task. No other main and interaction effects were significant.

**Table 1: Mean (SD) of post-task measures according to task complexity (\* $p < .05$ ).**

category	item	simple	complex	$F(1, 22)$	$p$ -value
workload	mental	2.58 (1.412)	2.88 (1.412)	1.286	.269
	temporal	2.54 (1.414)	3.38 (1.135)	5.670	.026*
	performance	3.96 (0.806)	3.63 (0.770)	2.514	.127
	effort	3.04 (1.268)	3.29 (1.197)	0.584	.453
satisfaction	enough info	3.96 (1.042)	3.46 (1.250)	2.760	.111
	time spent	3.92 (1.100)	3.12 (1.154)	5.954	.023*
	info amount	3.83 (1.167)	3.42 (1.176)	1.916	.180
	info quality	3.79 (1.179)	3.50 (1.180)	1.469	.238
	strategies	4.17 (0.761)	3.83 (0.816)	2.200	.152

**Effects on Search Behaviors (RQ2):** In RQ2, we investigate the effects of working memory and task complexity on participants’ search behaviors. As shown in Table 2, mixed-effects ANOVAs found a significant main effect of working memory on all four behavioral measures: (1) number of queries, (2) number of clicks, (3) avg. time to the first SERP click, and (4) avg. time between search actions. High-WM participants performed more actions and searched at a faster pace than low-WM participants.

**Table 2: Mean (SD) of behavioral measures according to working memory group (\* $p < .05$ ).**

measure	low-WM	high-WM	$F(1, 22)$	$p$ -value
queries	5.33 (3.595)	7.71 (4.298)	5.479	.029*
clicks	7.79 (5.013)	11.75 (5.252)	7.257	.013*
time to 1st click	12.58 (11.716)	6.98 (2.824)	5.063	.035*
time between events	56.67 (31.857)	38.47 (14.962)	5.355	.030*

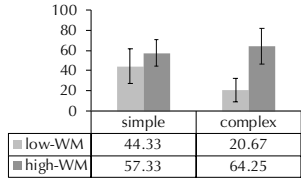
**Effects on Search Outcomes (RQ3):** In RQ3, we investigate the effects of working memory and task complexity on participants’ search outcomes. As shown in Table 3, mixed-effects ANOVAs found the following effects. First, working memory had a significant main effect on the length of participants’ justifications. High-WM participants produced significantly longer justifications than low-WM participants.

Second, task complexity had a significant main effect on the extent to which participants completed the corresponding  $n \times n$  table. Simple tasks had higher levels of *coverage* than complex tasks. In other words, participants were less successful in adding information to every cell in the table during more complex tasks.

Finally, working memory and task complexity had a significant interaction effect on the length of participants’ justifications ( $F(1, 22) = 5.713, p = 0.026$ ). As shown in Figure 1, high-WM participants produced longer justifications for both simple and complex tasks. However, the difference was *greater* for complex tasks.

**Table 3: Mean (SD) of search outcome measures according to working memory group and task complexity (\* $p < .05$ ).**

working memory (WM)	low-WM	high-WM	$F(1, 22)$	$p$ -value
coverage	0.93 (0.11)	0.96 (0.10)	0.882	0.358
justification length	32.50 (29.21)	60.79 (28.04)	9.079	.006*
task complexity	simple	complex	$F(1, 22)$	$p$ -value
coverage	1.00 (0.00)	0.89 (0.14)	12.942	.002*
justification length	50.83 (28.42)	42.46 (34.83)	1.714	.204



**Figure 1: Means and 95% confidence intervals of justification lengths (words) according to working memory group and task complexity.**

## 5 DISCUSSION AND CONCLUSION

In this study, we examined the effects of working memory and task complexity on participants’ post-task perceptions (RQ1), search behaviors (RQ2), and search outcomes (RQ3). In this section, we summarize our findings, compare them to those from prior studies, and discuss their implications.

**RQ1:** Task complexity had an effect on temporal demand and satisfaction with the time spent on the task. As expected, participants reported greater temporal demand and lower satisfaction for complex versus simple tasks. This serves as a manipulation check that our complex tasks required greater effort (as measured by temporal demand). Interestingly, working memory did not have a significant effect on any of our RQ1 measures.

**RQ2:** Working memory had a significant effect on several behavioral measures. Participants in the high working memory group performed more search actions (e.g., queries and clicks) and worked at a faster pace (e.g., took less time between subsequent actions) than participants in the low working memory group.

**RQ3:** In terms of search outcomes, we found three important trends. First, task complexity had an effect on participants ability to fully complete the  $n \times n$  table of items and dimensions. As expected, participants were less able to complete the table for complex versus simple tasks. This serves as a second sanity check that our manipulation of adding more items/dimensions increased the task’s complexity. Second, working memory had an effect on participants’ justification length. Participants in the high working memory group produced longer justifications than participants in the low working memory group. Finally, task complexity and working memory had a significant effect on participants’ justification length. High-WM participants produced *even longer* justifications than low-WM participants for complex tasks versus simple tasks.

Participants in our study were asked to write justifications based on information found regarding items and dimensions. In this respect, we treated longer justifications as better outcomes (i.e., more depth and/or breadth). Our RQ2 and RQ3 results suggest that high-WM participants exerted more effort (i.e., performed more search actions) and achieved better outcomes (i.e., produced longer justifications) than low-WM participants. Furthermore, in terms of the

justification length, this trend was more pronounced for complex tasks. Specifically, for complex tasks, high-WM participants produced justifications more than twice as long as those from low-WM participants. Interestingly, low-WM participants produced shorter justifications for complex versus simple tasks.

Our results suggest that low-WM participants were more likely to engage in satisficing behaviors, especially during more complex tasks. These results are largely consistent with prior studies, which found that low-WM participants were more likely to satisfice in challenging situations [16–18]. Also, our findings resonate with prior studies on the search behaviors of dyslexic users [14, 25]. In those studies, dyslexic participants were less “selective” than non-dyslexic participants in producing relevance judgments, possibly because they were less able to apply specific decision criteria and more likely to satisfice.

Interestingly, our RQ1 results are *inconsistent* with our RQ2 and RQ3 results. While high-WM participants exerted more effort (RQ2) and had better outcomes (RQ3), both participant groups reported *similar* levels of workload and satisfaction. In terms of workload, these results suggest that high-WM participants exerted more effort *without* perceiving greater levels of workload than low-WM participants. In terms of satisfaction, these results suggest that low-WM participants may have had a *lower* threshold for satisfaction with their performance. In other words, while low-WM participants exerted less effort and produced weaker justifications, they reported comparable levels of satisfaction as high-WM participants.

**Implications:** Our results have implications for designing systems that provide support and scaffolding for searchers with different working memory abilities. First and foremost, our results suggest that working memory plays an important role in performing comparative tasks, which are fairly common in everyday life (e.g., shopping). In our study, high-WM participants exerted more search effort and had better outcomes than low-WM participants. Future work should develop tools to support users (especially low-WM users) with comparative search tasks. Second, our results suggest that not all comparative tasks are equal. The number of items/dimensions being considered can impact the task’s complexity. Our results suggest that low-WM users may need even *more* support with complex comparative tasks in order to achieve better outcomes. Finally, our results suggest that low-WM users may have lower expectations about task outcomes and may therefore not *perceive* a need for support. While there were differences in performance between low- and high-WM participants, both groups reported similar levels of satisfaction. In this respect, low-WM users may benefit more from support systems that are readily available, visually salient, and possibly proactive.

**Future work:** Further research is needed to better understand the exact functions of working memory that are most crucial during the search process (e.g., planning, regulating, inferring relations between concepts and ideas, recognizing/addressing knowledge gaps, etc.). A deeper understanding of these functions may inform the design of tools to provide support and scaffolding for searchers with varying levels of working memory ability.

**Acknowledgements:** This work was supported in part by NSF grants IIS-1552587 and IIS-1451668. Any opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF.

## REFERENCES

- [1] National Aeronautics and Space Administration Human Systems Integration Division. 2016. TLX NASA Ames. (2016).
- [2] Lorin W. Anderson and David R. Krathwohl. 2001. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*.
- [3] Jaime Arguello. 2014. Predicting Search Task Difficulty. In *ECIR*. Springer, 88–99.
- [4] Alan Baddeley. 2010. Working memory. *Current biology* 20, 4 (2010), R136–R140.
- [5] Pierre Barrouillet and Jean-Francois Lecas. 1999. Mental models in conditional reasoning and working memory. *Thinking & Reasoning* 5, 4 (1999), 289–302.
- [6] Harald Beneventi, Finn Egil Tønnessen, and Lars Ersland. 2009. Dyslexic children show short-term memory deficits in phonological storage and serial rehearsal: an fMRI study. *International Journal of Neuroscience* 119, 11 (2009), 2017–2043.
- [7] Virginia W Berninger, Wendy Raskind, Todd Richards, Robert Abbott, and Pat Stock. 2008. A multidisciplinary approach to understanding developmental dyslexia within working-memory architecture: Genotypes, phenotypes, brain, and instruction. *Developmental neuropsychology* 33, 6 (2008), 707–744.
- [8] Kathy Brennan, Diane Kelly, and Jaime Arguello. 2014. The effect of cognitive abilities on information search for tasks of varying levels of complexity. In *III-X*. ACM, 165–174.
- [9] Donald J Campbell. 1988. Task complexity: A review and analysis. *Academy of management review* 13, 1 (1988), 40–52.
- [10] Robert Capra, Jaime Arguello, Anita Crescenzi, and Emily Vardell. 2015. Differences in the use of search assistance for tasks of varying complexity. In *SIGIR*. ACM, 23–32.
- [11] Ravit Cohen-Mimran and Shimon Sapir. 2007. Deficits in working memory in young adults with reading disabilities. *Journal of communication disorders* 40, 2 (2007), 168–183.
- [12] Meredyth Daneman and Patricia A Carpenter. 1980. Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior* 19, 4 (1980), 450–466.
- [13] Adele Diamond. 2013. Executive functions. *Annual review of psychology* 64 (2013), 135–168.
- [14] Adam Fourney, Meredith Ringel Morris, Abdullah Ali, and Laura Vonessen. 2018. Assessing the Readability of Web Search Results for Searchers with Dyslexia. In *SIGIR*. ACM.
- [15] Greg Francis, Angie MacKewn, and Danalee Goldthwaite. 2004. *CogLab on a CD*. Wadsworth Publishing Company.
- [16] Jacek Gwizdka. 2009. What a difference a tag cloud makes: effects of tasks and cognitive abilities on search results interface use. *Information Research: An International Electronic Journal* 14, 4 (2009), n4.
- [17] Jacek Gwizdka. 2013. Effects of Working Memory Capacity on Users' Search Effort. In *MIDI*. ACM, 11:1–11:8.
- [18] Jacek Gwizdka. 2017. I Can and So I Search More: Effects Of Memory Span On Search Behavior. In *CHIIR*. ACM, 341–344.
- [19] David Z. Hambrick and Randall W. Engle. 2003. *The Role of Working Memory in Problem Solving*. Cambridge University Press. 176–206 pages.
- [20] Xiao Hu and Noriko Kando. 2017. Task complexity and difficulty in music information retrieval. *JASIST* 68, 7 (2017), 1711–1723.
- [21] Bernard J Jansen, Danielle Booth, and Brian Smith. 2009. Using the taxonomy of cognitive learning to model online searching. *Information Processing & Management* 45, 6 (2009), 643–663.
- [22] 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 *ICTIR*. ACM, 101–110.
- [23] Patrick C Kyllonen and Raymond E Christal. 1990. Reasoning ability is (little more than) working-memory capacity?! *Intelligence* 14, 4 (1990), 389–433.
- [24] Cindy Lustig, Cynthia P May, and Lynn Hasher. 2001. Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General* 130, 2 (2001), 199.
- [25] Andrew MacFarlane, A Albair, CR Marshall, and George Buchanan. 2012. Phonological working memory impacts on information searching: an investigation of dyslexia. In *III-X*. ACM, 27–34.
- [26] James Smith-Spark, John Fisk, Angela Fawcett, and Roderick Nicolson. 2003. Investigating the central executive in adult dyslexics: Evidence from phonological and visuospatial working memory performance. *European Journal of Cognitive Psychology* 15, 4 (2003), 567–587.
- [27] Marilyn L Turner and Randall W Engle. 1989. Is working memory capacity task dependent? *Journal of memory and language* 28, 2 (1989), 127–154.
- [28] Lauren Turpin, Diane Kelly, and Jaime Arguello. 2016. To Blend or Not to Blend?: Perceptual Speed, Visual Memory and Aggregated Search. In *SGIR*. ACM, 1021–1024.
- [29] Barbara Wildemuth, Luanne Freund, and Elaine G. Toms. 2014. Untangling search task complexity and difficulty in the context of interactive information retrieval studies. *Journal of Documentation* 70, 6 (2014), 1118–1140.
- [30] Barbara Wildemuth, Diane Kelly, Emma Boettcher, Erin Moore, and Gergana Dimitrova. 2018. Examining the impact of domain and cognitive complexity on query formulation and reformulation. *IP&M* 54, 3 (2018), 433 – 450.
- [31] Jennifer Wiley and Andrew F Jarosz. 2012. How working memory capacity affects problem solving. In *Psychology of learning and motivation*. Vol. 56. Elsevier, 185–227.
- [32] Wan-Ching Wu and Diane Kelly. 2014. Online search stopping behaviors: An investigation of query abandonment and task stopping. *ASIST* 51, 1 (2014), 1–10.