The Effects of Thinking Aloud on Participants during Search and Sensemaking

Bogeum Choi
University of Illinois Urbana-Champaign
Illinois, USA
bogeumc@illinois.edu

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

Abstract

During a think-aloud study, participants verbalize their thoughts as they complete a specified task. Think-aloud comments provide insights into what a participant is doing and experiencing in the moment. Interactive information retrieval studies have used think-aloud protocols to explore different research questions. However, think-aloud protocols can also influence participants. We report on a qualitative analysis of data collected during a think-aloud study. Participants were asked to learn about a complex topic by searching online and taking notes. After the task, participants were asked whether and how thinking aloud influenced their approach to the task. Open-ended responses revealed 21 (positive and negative) ways in which thinking aloud influenced participants. Additionally, during the study, we measured participants' working memory (WM) capacity. We did not find that thinking aloud had different influences for low- vs. high-WM participants.

CCS Concepts

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

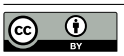
Keywords

Think-aloud protocol, search and sensemaking, working memory

ACM Reference Format:
Bogeum Choi and Jaime Arguello. 2026. The Effects of Thinking Aloud on Participants during Search and Sensemaking. In *2026 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '26), March 22–26, 2026, Seattle, WA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3786304.3787883

1 Introduction

A think-aloud (TA) protocol is a research method where participants verbalize their thoughts as they complete a specified task. TA comments provide insights about what a study participant is experiencing in the moment—goals, motivations, focus, (meta)cognitive processes, challenges, perceptions, and/or feelings. TA protocols are commonly used in interactive information retrieval (IIR) studies. IIR studies have used TA protocols to understand how searchers: (1) select search tactics [40], (2) judge relevance [20], (3) judge credibility [23, 24], and (4) decompose learning objectives [38].



CHIIR '26, Seattle, WA, USA

ACM ISBN 979-8-4007-2414-5/2026/03
https://doi.org/10.1145/3786304.3787883

While TA protocols can provide IIR researchers with a treasure trove of insights, they also have the potential to influence behavior. Prior research has found both negative and positive effects. For example, some studies have found that it increases cognitive load [35] and others have argued that thinking aloud can enhance metacognitive engagement and reflection [1, 6, 30, 33].

We report on a qualitative analysis of data collected (but *not* analyzed) during a previous think-aloud study [3]. During the study, 44 participants completed a single search and sensemaking (SSM) task. Participants were asked to learn about a complex topic by searching online and taking notes. After the task, a post-task questionnaire included two questions about the TA protocol. First, participants were asked if the TA protocol influenced their approach to the task. Second, they were asked to elaborate on their answer to the first question. For this paper, we report on a qualitative analysis of open-ended responses to the second question.

TA protocols may influence participants differently depending on their individual characteristics—cognitive abilities, personality traits, etc. During the study, we measured participants' working memory (WM) capacity—their ability to store and manipulate information in short-term memory. Participants were binned into low- and high-WM groups (22 per group). Thinking aloud is the process of verbalizing thoughts in working memory. Therefore, in this paper, we also explore whether thinking aloud had a *different* influence for low- vs. high-WM participants. To summarize, we explore two research questions:

- **RQ1:** How did participants perceive to have been impacted by thinking aloud during the SSM task?
- **RQ2:** Did WM affect participants' perceived impact of thinking aloud during the SSM task?

2 Background

The Think-Aloud Protocol: A think-aloud (TA) protocol is a data elicitation method used to gain access to people's cognitive processes such as their thoughts, goals, motivations, and decisions. It asks participants to verbalize whatever comes to mind while completing a task. TA protocols have been widely used in studies of decision making [13], reading [21, 33], and learning [7–9, 32]. In information science, TA protocols have been used to study search tactics [40], relevance criteria [20], credibility assessments [23, 24], and goal-setting behaviors during complex search tasks [38]. In the health domain, TA protocols have been used to study how people search for health information [27], judge the efficacy of medical treatments [16], and self-diagnose [25]. When used alongside screen recordings, think-aloud comments can help clarify participants' goals, reading behaviors, and attention [41].

Ericsson and Simon [14] defined three levels of verbalization based on the cognitive processes involved. Level 1 involves saying

aloud what one would normally say silently. Level 1 verbalizations may include reading text aloud or expressing emotions without filtering. Level 2 involves translating non-verbal information into verbal form. Level 2 verbalizations may happen when someone describes their actions, an image, or an emerging pattern. Finally, level 3 requires even greater information processing. Level 3 verbalizations may happen when someone is asked to justify their actions or reasoning, often prompted by the moderator asking: "Why did you do that?". Higher-level verbalizations are more likely to influence behaviors. Our study involved Level 1 and 2 verbalizations. The moderator simply prompted participants to "please keep taking" if they were silent for too long.

Effects of Think-Aloud Protocol on Task Performance: The TA protocol, while valuable for revealing participants' cognitive processes, has raised concerns about its potential effects on task performance. A commonly cited limitation is that verbalizing thoughts can increase cognitive load [35]. Thinking aloud may also alter the very cognitive processes under investigation by causing more deliberate, analytical modes of thinking.

Empirical findings are mixed. Some studies have reported that TA protocols can increase cognitive load [29] and alter behaviors such as web page navigation and scrolling behaviors [29], link traversal behaviors [28], and task success rates [28]. However, these effects have typically been found for TA protocols that go beyond Level 1-2 verbalizations (e.g., "Tell me what you like, dislike, and find confusing" vs. "Act like you're alone in the room speaking to yourself"). Therefore, the effects of thinking aloud on behaviors and performance likely depend on factors such as instructions, prompts, and task complexity. Finally, some researchers have argued that thinking aloud may have positive impacts on task performance by enhancing metacognitive engagement and reflection [1, 6, 30, 33].

Working Memory: Working memory (WM) is the cognitive system responsible for the short-term storage and manipulation of information, even when not perceptually present. WM plays an important role in cognitive tasks such as reading [2, 10, 12], comprehension [31, 42], logical reasoning [11], and attentional control [26]. Several studies have explored the influence of WM during complex search tasks. Prior work suggests that high-WM searchers exert more effort, implying that low-WM searchers are more likely to satisfice [4, 5, 18, 19]. Sharit et al. [36] found that WM predicted search task performance among older adults. Pardi et al. [34] found that high-WM participants recalled a wider range of concepts after searching to learn about a topic. In this paper, we examine how thinking aloud during a complex search task may affect participants with high vs. low WM capacity. Think-aloud comments capture thoughts that are currently active in WM [39]. Therefore, it is reasonable to question whether and how WM capacity influences the effects of thinking aloud in this context.

3 Methods

To investigate RQ1 & RQ2, we analyzed data collected as part of a prior study that employed a TA protocol [3]. In this section, we summarize the study and results. We focus on details relevant to RQ1 & RQ2. Then, we describe our qualitative analysis of participants' open-ended responses about the perceived impacts of thinking aloud during the main task of the study.

3.1 Summary of Study and Results

The study investigated the role of working memory (WM) during a complex search and sensemaking (SSM) task. Forty-four participants were binned into a high- and low-WM group (22 per group) based on their scores on a psychometric test called the Operation Span (OSPAN) task [15]. Participants (aged 18-65) were recruited through an opt-in mailing list of employees at our university and received US$40 for participating.

During the study, participants were asked to work on a SSM task (Section 3.2) for 30 minutes. After the task, participants completed a learning assessment that asked them to describe everything they learned. Responses were manually scored, and participants in the high-WM group had significantly better learning outcomes. Additionally, participants completed questionnaires about their perceptions before and after the task. There were no significant differences in pre- and post-task perceptions between groups.

Participants were asked to think aloud during the SSM task. Participants were instructed to "narrate their stream of consciousness". The study moderator was present in the room and prompted participants to "please keep talking" if they were silent for too long. Video recordings of participants' screen activity and think-aloud comments were analyzed to measure engagement in specific SSM and (meta)cognitive activities. Results found that high-WM participants were more likely to engage in activities such as progress monitoring, drawing connections, and evaluating semantic fit, among others.

Relevant to the current paper, during the post-task questionnaire, participants were asked two questions about the TA protocol. First, they indicated their level of agreement with the statement: "Having to think aloud during the search task greatly impacted my approach to the task." using a 7-point scale (1 = strongly disagree to 7 = strongly agree). Second, participants were asked: "In a few sentences, please describe your response to the previous question."

3.2 Search & Sensemaking Task

We define a search and sensemaking (SSM) task as one that requires both information searching and sensemaking to achieve the task goal, which in our case was learning about a complex topic. Participants completed the following SSM task:

Scenario: You recently attended a lecture about the gut-brain connection. The speaker explained the relationship between the gut microbiome and one's physical and mental health. After the lecture, you realized there is much more to learn about the connection between your digestive system and your overall well-being.

Objective: Try to find out and learn about the topic of the gut microbiome and an individual's physical and mental health. Potential sub-topics you can explore include but are not limited to: What is the notion of "gut-brain connection"? Through what mechanisms do gut microbiota influence one's physical and mental health? What factors can influence gut microbiota? What are some science-backed ways to improve your gut health?

3.3 Qualitative Analysis

As mentioned in Section 3.1, during the post-task questionnaire, participants indicated the extent to which thinking aloud impacted their approach to the task and elaborated on their response. The second open-ended question aimed to capture participants' perceptions of the TA protocol and its impact during the SSM task.

Our qualitative analysis of open-ended responses involved two researchers: R1 and R2. First, R1 analyzed all 44 participants' responses with the goal of identifying perceived impacts of thinking aloud. Based on this, R1 developed a codebook comprising codes, definitions, and example quotes. Then, R1 and R2 met to discuss and refine the codebook. Ultimately, we developed 21 codes. We anticipated some codes to be very infrequent in the data. However, we were mostly interested in the *variety* of ways in which thinking aloud impacted participants. Additionally, codes were organized into positive and negative groups, reflecting participants' mentions of both positive effects (e.g., helped with retention) and negative effects (e.g., slowed me down).

Next, using the final codebook, R1 and R2 independently coded 100% of the data. Codes were designed to be *not* mutually exclusive—responses could be assigned multiple codes. After independently coding the data, we measured agreement using Cohen's κ. Ten codes had 'almost perfect' agreement ($\kappa > 0.80$); 10 codes 'substantial' agreement ($0.80 \geq \kappa > 0.60$); and 1 code 'moderate' agreement ($0.60 \geq \kappa > 0.40$) [22]. Finally, R1 and R2 met to resolve all disagreements and finalize the assignment of codes.

4 RQ1 Results: Effects of Thinking Aloud

We found 21 ways in which thinking aloud influenced participants. For each code, we provide three values. N is number of participants (out of 44) whose response was assigned the code. L and H is the number of participants from the low- and high-WM group whose response was assigned the code (out of 22 per group).

Negative Codes: Participants commented on seven different ways in which thinking aloud had negative impacts.

- **Extra Time (**$N = 7$, $L = 2$, $H = 5$**):** Participants mentioned that thinking aloud required additional time or slowed them down. P1 mentioned: "I had to use time narrating what I'd decided to do after deciding to do it." P6 mentioned: " I felt like speaking aloud slowed down my process of learning." P3 wrote: "I would normally move between thoughts quicker."
- **Fatigue (**$N = 1$, $L = 0$, $H = 1$**):** One participant mentioned that thinking aloud made them feel tired. P8 mentioned: "But I also got tired much faster with such great amounts of concentration."
- **Divided Attention (**$N = 8$, $L = 4$, $H = 4$**):** Participants wrote that their attention was split between thinking aloud and other tasks (e.g., searching, reading, and note-taking). P21 mentioned: "I was a bit distracted from the reading sometimes because I was thinking about how frequently I had talked and whether I had to talk again soon." P18 wrote: "It was harder to process the direction I wanted to go with my research on the topic when I had to also concentrate on thinking out loud."
- **Disrupted Focus (**$N = 3$, $L = 0$, $H = 3$**):** Participants wrote that having to pause and verbalize their thoughts disrupted their "flow". P5 wrote: "My own voice interrupted my thought process."
- **Disorienting (**$N = 1$, $L = 0$, $H = 1$**):** One participant mentioned that thinking aloud made it harder to stay focused on task-relevant activities. P5 mentioned: "... I was more likely to go on tangents in following my trains of thought out loud to places that were not necessarily relevant to the task itself."
- **Observer Effect (**$N = 3$, $L = 1$, $H = 2$**):** Participants mentioned that thinking aloud made them feel self-conscious. P21 noted: "It [...] led to me taking the research more seriously so I didn't sound stupid." P29 wrote: "I was more self-conscious [...] I wondered what the observer was thinking. Did they think my thought process was competent enough for a PhD student?"
- **Unfamiliar/Unnatural (**$N = 10$, $L = 4$, $H = 6$**):** Participants mentioned being unfamiliar with verbalizing their thoughts while working on the task. P11 mentioned: "I suppose I've conditioned myself not to think out loud." P23 wrote: "It was just an extra thing to do and felt unnatural."

Positive Codes: Participants commented on 14 different ways in which thinking aloud had positive impacts.

- **Focus (**$N = 5$, $L = 3$, $H = 2$**):** Participants mentioned that thinking aloud helped them stay focused. P8 mentioned: "Thinking out loud let me zone out and fully concentrate on the task."
- **Comprehension (**$N = 3$, $L = 2$, $H = 1$**):** Participants mentioned that thinking aloud helped them understand things better or faster. P15 wrote: "I found myself catching when I was unconsciously making connections to other info (to which I'd go back and rethink)." P20 wrote: "It helps me to read and understand what it means faster."
- **Checking Comprehension (**$N = 5$, $L = 4$, $H = 1$**):** Related to the previous code, participants mentioned that thinking aloud prompted them to *assess* their understanding of information. P3 mentioned: "Speaking out loud helped me [...] test my own understanding." P4 wrote: "reading aloud helps me truly check whether I understand the terms or concepts I am reading about."
- **Recognizing Significance (**$N = 3$, $L = 3$, $H = 0$**):** Participants mentioned that thinking aloud sometimes helped them realize the importance of what they were reading or thinking. P19 noted: "It made me realize when thoughts I was having could be considered significant." P10 wrote: "Saying my findings aloud made me reconsider whether they were useful/important."
- **Increased Awareness of Thought Process (**$N = 6$, $L = 3$, $H = 3$**):** Participants mentioned that thinking aloud made them more aware of what and how they were thinking. P16 mentioned: "[It] made me more aware of my thought process." P30 noted: "[It] made me think about what I was actually looking at."
- **Increased Awareness of Goals (**$N = 3$, $L = 1$, $H = 2$**):** Participants mentioned that thinking aloud increased their awareness of their goals or lack thereof. P42 mentioned: "It gave clarity to what goal I had in mind." P30 noted: "It made me more goal-oriented because I kept realizing that I didn't have goals in mind."
- **Increased Awareness of Progress (**$N = 3$, $L = 0$, $H = 3$**):** Thinking aloud helped participants keep track of their progress. P20 noted: "[It] helped me to locate the progress that I had already made." P27 wrote: "[It] made me think about whether I had enough information about the topic before moving on."
- **Increased Awareness of Performance (**$N = 1$, $L = 1$, $H = 0$**):** One participant mentioned that thinking aloud made them aware of their performance during certain activities. P7 mentioned: "I was more aware of the efficiency of my searches."
- **Stay on Track (**$N = 4$, $L = 3$, $H = 1$**):** Participants mentioned that thinking aloud helped them stay on course and avoid going down tangential paths that did not align with the task goals. P4 noted: "Thinking aloud does help me not get side-tracked." P7 wrote: "I was less likely to go on a tangent."
- **Organized (**$N = 5$, $L = 2$, $H = 3$**):** Thinking aloud helped participants structure or organize their thoughts. P34 noted: "Having

to think aloud did help me to outline my process." P40 wrote: "I could organize my thoughts by speaking them aloud."

- **Intentional ($N = 5$, $L = 4$, $H = 1$):** Thinking aloud increased participants' overall awareness, which in turn led them to be more intentional and to reconsider the reasons behind their actions. P7 noted: "I was more aware of the meaning behind every choice I made." P10 noted: "Saying my search approach [...] made me second guess why I was doing it that way."
- **Follow-through ($N = 1$, $L = 0$, $H = 1$):** One participant mentioned that verbalizing their plans made them feel more accountable and committed. P21 wrote: "[Thinking aloud] made me choose a path [and] then stick to it."
- **Help with Retention ($N = 5$, $L = 2$, $H = 3$):** Participants mentioned that thinking aloud helped them remember what they read and learned. P38 noted: "Just the effort to verbalize what I was learning helped me remember it." P40 noted: "I just mumbled aloud any sections of the readings that were important so that I could better remember them."
- **Help with Writing ($N = 1$, $L = 1$, $H = 0$):** One participant mentioned that thinking aloud helped them with writing their notes. P41 wrote: "I was able to write more effectively and concisely."

5 RQ2 Results: Thinking Aloud & WM

In RQ2, we investigate whether WM impacted the effects of thinking aloud on participants. First, we considered the effects of WM on participants' Likert scale response to the prompt: "Having to think aloud during the task greatly impacted my approach to the task." A Wilcoxon rank sum test did not find significant differences betweeen high- and low-WM groups.

Second, we investigated whether responses from one group were more (or less) likely to be associated with each code described in Section 4. For each code, we conducted a logistic regression analysis using WM group as the independent variable and the presence/absence of the code as the dependent variable. WM group did not have a significant effect on the likelihood of any code.

Finally, we investigated whether WM had an effect on the number of positive/negative codes associated with each response. To this end, we conducted a Poisson regression analysis using WM group as the independent variable and the number of positive/negative codes associated with the response as the dependent variable. Responses from the low-WM group had a greater number of positive codes (low-WM = 26, high-WM = 22). However, this effect was not significant. Correspondingly, responses from the high-WM group had a greater number of negative codes (high-WM = 22, low-WM = 11). However, this effect was only *marginally* significant ($p = .061$). Taken as a whole, our RQ2 results suggest that WM did not impact the effects of thinking aloud on participants.

6 Discussion & Conclusion

In this study, we examined how participants perceived the impacts of the TA protocol during a task that requires both searching and sensemaking to learn about a complex topic. Overall, participants reported a wide range of positive and negative effects. Interestingly, some effects directly contradicted each other. For example, some participants reported that thinking aloud interrupted their thought process (**disrupted focus**) and others said that it helped them concentrate (**focus**). Similarly, some participants said that thinking aloud made them prone to tangents (**disorienting**) and others said it helped them stay aligned with the task goals (**stay on track**). These contradicting accounts suggest that the perceived effects of the TA protocol are not uniform, but may instead vary across individuals and contexts. Future research should explore how factors such as cognitive style, reading mode preference [37], or task characteristics affect the ways in which TA protocols influence the learning experience during search.

Our RQ2 results largely found that WM capacity did not influence the perceived impacts of the TA protocol. This finding may seem counterintuitive. Given that thinking aloud is considered to burden working memory resources, one might expect lower WM searchers to find it more disruptive. Our results suggest that the relationship between WM and the impacts of the TA protocol is more complex.

While only *marginally* significant ($p = .061$), high-WM participants described more negative ways in which thinking aloud influenced them. That is, their open-ended responses had a greater variety of negative codes. Some prior studies have also found that thinking aloud can have more detrimental effects for high-WM participants. For example, Goo [17] found that high-WM learners were more negatively impacted by thinking aloud during a language-learning task than low-WM learners.

Notably, several positive codes centered on increased metacognitive awareness (i.e., of **thought processes**, **goals**, **progress**, and **performance**). This pattern is consistent with prior work that has characterized thinking aloud as a strategy for enhancing metacognitive processes [1, 6]. These findings suggest that a TA protocol may, in some cases, serve as a scaffold that makes activities such as monitoring and checking comprehension more explicit, thereby facilitating learning for low-WM individuals. For high-WM individuals, however, these processes may already occur automatically, making verbalization feel redundant or disruptive.

Limitations: This study has two important limitations. First, it draws on self-reported data, which has two caveats. One is that participants may only report experiences that are salient at the moment, meaning the data may not capture all effects they experienced. The second is that survey data reflects perceptions rather than actual outcomes. For example, feeling more focused while thinking aloud does not necessarily imply improved concentration. Second, the sample size was relatively small, which may have limited our ability to detect a significant effect of WM.

Conclusion: Our study makes several contributions. First, we presented a wide range of ways in which participants can be both positively and negatively influenced by thinking aloud during an IIR study. Given the utility and common use of TA protocols in IIR studies, it is important to understand whether and how thinking aloud influences participants' perceptions. Second, we show that thinking aloud can be experienced in *contradictory* ways. This suggests that researchers should avoid making general assumptions about the effects of TA protocols on participants' performance or experience, warranting further investigation into individual differences in reactivity. Third, to our knowledge, this is the first study to examine the relationship between WM and thinking aloud in the context of learning during search. Our results found no significant differences in perceptions between high- and low-WM participants. Together, these contributions enhance our understanding of TA protocols as a methodological tool.

References

[1] Cathy Collins Block and Susan E Israel. 2004. The ABCs of performing highly effective think-alouds. *The Reading Teacher* 58, 2 (2004), 154–167.

[2] Judy Cantor, Randall W Engle, and George Hamilton. 1991. Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence* 15, 2 (1991), 229–246.

[3] Bogeum Choi and Jaime Arguello. 2025. The Effects of Working Memory during a Search and Sensemaking Task. In *Proceedings of the 2025 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '25)*. Association for Computing Machinery, New York, NY, USA, 100–111. https://doi.org/10.1145/3698204.3716449

[4] Bogeum Choi, Jaime Arguello, Robert Capra, and Austin R. Ward. 2023. The influences of a knowledge representation tool on searchers with varying cognitive abilities. *ACM Transactions on Information Systems* 41, 1, Article 18 (2023), 35 pages.

[5] Bogeum Choi, Robert Capra, and Jaime Arguello. 2019. The effects of working memory during search tasks of varying complexity. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. ACM, New York, NY, USA, 261–265.

[6] Julie Coiro. 2011. Talking about reading as thinking: Modeling the hidden complexities of online reading comprehension. *Theory Into Practice* 50, 2 (2011), 107–115.

[7] Deborah Cotton and Karen Gresty. 2006. Reflecting on the think-aloud method for evaluating e-learning. *British Journal of Educational Technology* 37, 1 (2006), 45–54.

[8] DRE Cotton and KA Gresty. 2007. The rhetoric and reality of e-learning: using the think-aloud method to evaluate an online resource. *Assessment & Evaluation in Higher Education* 32, 5 (2007), 583–600.

[9] Michael S Crowther, Chris C Keller, and Gregory L Waddoups. 2004. Improving the quality and effectiveness of computer-mediated instruction through usability evaluations. *British Journal of Educational Technology* 35, 3 (2004), 289–303.

[10] 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.

[11] Wim De Neys, Walter Schaeken, and Géry d'Ydewalle. 2005. Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning* 11, 4 (2005), 349–381.

[12] Randall W Engle, John K Nations, and Judy Cantor. 1990. Is "working memory capacity" just another name for word knowledge? *Journal of Educational Psychology* 82, 4 (1990), 799–804.

[13] K Anders Ericsson and Jared H Moxley. 2011. Thinking aloud protocols: Concurrent verbalizations of thinking during performance on tasks involving decision making. *A handbook of process tracing methods for decision research* (2011), 89–114.

[14] K Anders Ericsson and Herbert A Simon. 1980. Verbal reports as data. *Psychological review* 87, 3 (1980), 215.

[15] G. Francis, I. Neath, A. MacKewn, and D. Goldthwaite. 2004. *CogLab on a CD*. Thomson/Wadworth. https://books.google.com/books?id=ikhqmEXx7AIC

[16] Amira Ghenai, Mark D Smucker, and Charles LA Clarke. 2020. A think-aloud study to understand factors affecting online health search. In *Proceedings of the 2020 conference on human information interaction and retrieval*. 273–282.

[17] Jaemyung Goo. 2010. Working memory and reactivity. *Language Learning* 60, 4 (2010), 712–752.

[18] Jacek Gwizdka. 2013. Effects of working memory capacity on users' search effort. In *Proceedings of the International Conference on Multimedia, Interaction, Design and Innovation*. ACM.

[19] Jacek Gwizdka. 2017. I can and so I search more: effects of memory span on search behavior. In *Proceedings of the 2017 Conference on Human Information Interaction and Retrieval*. New York, NY, USA, 341–344.

[20] Sandra G Hirsh. 1999. Children's relevance criteria and information seeking on electronic resources. *Journal of the American Society for information Science* 50, 14 (1999), 1265–1283.

[21] Joan E Hughes, B Wai-Ling Packard, and P David Pearson. 1997. Reading classroom explorer: Visiting classrooms via hypermedia. In *YEARBOOK-NATIONAL READING CONFERENCE*, Vol. 46. THE NATIONAL READING CONFERENCE, INC., 494–506.

[22] J. R. Landis and G. G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 1 (1977), 159–174.

[23] Teun Lucassen, Rienco Muilwijk, Matthijs L Noordzij, and Jan Maarten Schraagen. 2013. Topic familiarity and information skills in online credibility evaluation. *Journal of the American Society for Information Science and Technology* 64, 2 (2013), 254–264.

[24] Teun Lucassen and Jan Maarten Schraagen. 2010. Trust in wikipedia: how users trust information from an unknown source. In *Proceedings of the 4th workshop on Information credibility*. 19–26.

[25] Tana M Luger, Thomas K Houston, and Jerry Suls. 2014. Older adult experience of online diagnosis: results from a scenario-based think-aloud protocol. *Journal of medical Internet research* 16, 1 (2014), e2924.

[26] Xiaoxiao Luo, Liwei Zhang, and Jin Wang. 2017. The benefits of working memory capacity on attentional control under pressure. *Frontiers in psychology* 8 (2017), 1105.

[27] Wendy Macias, Mina Lee, and Nicole Cunningham. 2018. Inside the mind of the online health information searcher using think-aloud protocol. *Health communication* 33, 12 (2018), 1482–1493.

[28] Sharon McDonald, Kenneth McGarry, and Leanne M Willis. 2013. Thinking-aloud about web navigation: the relationship between think-aloud instructions, task difficulty and performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 57. SAGE Publications Sage CA: Los Angeles, CA, 2037–2041.

[29] Sharon McDonald and Helen Petrie. 2013. The effect of global instructions on think-aloud testing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2941–2944.

[30] Regina G McKeown and James L Gentilucci. 2007. Think-aloud strategy: Metacognitive development and monitoring comprehension in the middle school second-language classroom. *Journal of Adolescent & Adult Literacy* 51, 2 (2007), 136–147.

[31] Jane Oakhill, Joanne Hartt, and Deborah Samols. 2005. Levels of Comprehension Monitoring and Working Memory in Good and Poor Comprehenders. *Reading and Writing* 18, 7 (2005), 657–686.

[32] Martin Oliver, Judith MacBean, Grainne Conole, and Jen Harvey. 2002. Using a toolkit to support the evaluation of learning. *Journal of Computer Assisted Learning* 18, 2 (2002), 199–208.

[33] Leslie Oster. 2001. Using the think-aloud for reading instruction. *The Reading Teacher* 55, 1 (2001), 64–69.

[34] Georg Pardi, Johannes von Hoyer, Peter Holtz, and Yvonne Kammerer. 2020. The Role of Cognitive Abilities and Time Spent on Texts and Videos in a Multimodal Searching as Learning Task. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. ACM, 378–382.

[35] Babette Park, Andreas Korbach, and Roland Brünken. 2020. Does thinking-aloud affect learning, visual information processing and cognitive load when learning with seductive details as expected from self-regulation perspective? *Computers in Human Behavior* 111 (2020), 106411. https://doi.org/10.1016/j.chb.2020.106411

[36] Joseph Sharit, Mario A. Hernández, Sara J. Czaja, and Peter Pirolli. 2008. Investigating the Roles of Knowledge and Cognitive Abilities in Older Adult Information Seeking on the Web. *ACM Transactions on Computer-Human Interaction* 15, 1 (2008).

[37] Ioannis Smyrnakis, Vassilios Andreadakis, Andriani Rina, Nadia Boufachrentin, and Ioannis M Aslanides. 2021. Silent versus reading out loud modes: An eye-tracking study. *Journal of eye movement research* 14, 2 (2021), 10–16910.

[38] Kelsey Urgo and Jaime Arguello. 2022. Understanding the "Pathway" Towards a Searcher's Learning Objective. *ACM Transactions of Information Systems* 40, 4, Article 77 (Jan. 2022), 43 pages. https://doi.org/10.1145/3495222

[39] Maarten W Van Someren, Yvonne F Barnard, Jacobijn AC Sandberg, et al. 1994. The think aloud method: a practical approach to modelling cognitive processes. *London: AcademicPress* 11, 6 (1994).

[40] Iris Xie and Soohyung Joo. 2012. Factors affecting the selection of search tactics: Tasks, knowledge, process, and systems. *Information Processing & Management* 48, 2 (2012), 254–270.

[41] Kirsty A Young. 2005. Direct from the source: the value of'think-aloud'data in understanding learning. *Journal of Educational Enquiry* (2005).

[42] Nicola Yuill, Jane Oakhill, and Alan Parkin. 1989. Working memory, comprehension ability and the resolution of text anomaly. *British journal of psychology* 80, 3 (1989), 351–361.