Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Enhancing Retrospective Think-Aloud Protocols with Eye-Gaze Data

Amit PAIKRAO*, Debarshi NATH, Ramkumar RAJENDRAN, & Ritayan MITRA
Centre for Educational Technology, Indian Institute of Technology Bombay, India
*amitpaikrao@iitb.ac.in

Abstract: Recent advances in cognitive research indicate potential in integrating eye-tracking patterns with traditional retrospective think-aloud protocols. However, the exploration of this approach in learning and education has not taken place yet. This study compares conventional video-stimulated retrospective think-aloud with an eye-gaze-enhanced retrospective think-aloud method, examining their effectiveness in eliciting process-level cognitive data and metacognitive awareness. Preliminary findings indicate integration of eye gaze with retrospective think-aloud enhances verbalization precision of cognitive process-level data, reduces attentional ambiguity, minimizes cognitive reconstruction artifacts, and enables the formulation of targeted probes. These outcomes, aligned with Flavell's metacognitive monitoring framework, highlight multimodal protocol analysis in capturing the dynamic interplay between visual attention, decision-making, and self-regulated learning.

Keywords: Think-aloud, video-stimulated retrospective think-aloud, eye-tracking, eye gaze enhanced retrospective think-aloud, cognitive process level data, metacognition.

1. Introduction and Literature Review

Insights into cognitive processes that take place in computer-based learning environments remain significantly challenging, as the complex interplay of their thoughts, decisions, and metacognitive reflections often remains latent. In such scenarios, think-aloud protocols have emerged as valuable tools that allow researchers to gain deeper insights by prompting learners to verbalize their experiences (Ericsson & Simon, 1980). However, reliance on memory and verbal articulation inherently introduces abstraction that may obscure the fine-grained dynamics (Ericsson & Simon, 1993). Primarily, there are two variants of think-aloud protocols: concurrent think-aloud (CTA) and retrospective think-aloud (RTA) (Ericsson & Simon, 1980). In CTA, participants verbalise their thoughts while simultaneously performing the task. This method provides immediate insight into information held in working memory (W. Rainey Johnson et al. 2023). On the other hand, the retrospective think-aloud (RTA) method involves participants reflecting on their actions, decisions, and reasoning without the cognitive burden of simultaneous task performance and verbalization (Ericsson & Simon, 1993). Nonetheless, RTA's efficacy depends upon participants' memory recall capabilities, with a temporal delay between task completion and reflection introducing potential memory decay effects. When RTA implementation is not proximal to task execution, authentic cognitive processes may weaken significantly. RTA is inherently vulnerable to various response biases. Participants frequently engage in post-hoc cognitive reconstruction rather than accurate recall, thereby introducing threats to internal validity and maturation effects that compromise methodological rigour (Noushad et al., 2023). Additionally, conventional RTA protocols may inadequately capture the dynamic complexity of cognitive mechanisms, particularly in contexts involving visual-attentional processes and graphical interface interaction (Johnson et al., 2023). Given the limitations of traditional RTA methods, integrating RTA with video stimulation, known as video-stimulated recall (VS-RTA) (Lyle, 2003; Alhadreti & Mayhew, 2018), helps to overcome some of these limitations. VS-RTA has been used in teaching, counselling, psychiatry/medical research, nursing, and counselling studies. Video-stimulated think-aloud involves replaying the video of the task interaction during the retrospective interviews (Paikrao

& Mitra, 2023). This approach effectively leverages visual records to cue memory and helps to facilitate detailed reflections on the learning and problem-solving process. However, it is also susceptible to rationalization and reconstruction of events.

The subsequent integration in RTA involves the use of eye gaze data. While eye gaze data information is extremely valuable on its own (Rajendran et al., 2018), eye-tracking integrations in RTA, specifically superimposition of eye gaze with video and gaze plot for webpages. The eye-gaze data is overlaid onto the recorded interaction video in such a method. The resulting merged video is then used to prompt retrospective interviews with participants to elicit responses (Olsen et al., 2010; Elbabour et al., 2017). We refer to this integration of eye-gaze data (including fixations, saccades, and eye-gaze patterns) with retrospective think-aloud as ET-RTA. This methodology has been primarily explored in usability testing to identify more specific usability problems (Dewi et al., 2023; Olsen et al., 2010; Elbabour et al., 2017; Ishammari et al., 2015; Alhadreti & Mayhew, 2018; Pathan et al., 2025). The application of this methodology in educational contexts is minimal, and its potential can only be realised when this approach is investigated in different learning contexts. Existing research has not been able to address how eye-tracking technologies can be used in conjunction with the think-aloud protocol and what advantages such approaches can offer over traditional video-stimulated retrospective think-aloud. Given this predominant focus on usability testing problems, there remains a compelling research imperative to explore this combined methodology in the context of learning to shed light on cognitive and metacognitive processes. This identified research gap has motivated us to compare the video-stimulated retrospective think-aloud (VS-RTA) method with the eye-gaze enhanced retrospective think-aloud (ET-RTA). The pertinent research question is: How eye tracking integration enhances *retrospective think-aloud methodology beyond traditional video stimulation?* It seeks to extend eye-tracking applications beyond traditional usability domain toward a more comprehensive understanding of cognitive processes.

2. Methodology

We employed a single-participant within-subjects design as a proof-of-concept investigation with a single higher education student from electrical engineering. The goal was to trace the epistemic frames of participants as they engage in the role-playing game “The Power to the People” (https://www.rhombicogames.com/game/power-to-the-people/). The task in this game was to design an electrical grid and maintain a continuous power supply to the residents of the city. The study comprised three phases: gameplay (16 minutes), VS-RTA (30 minutes), and ET-RTA (42 minutes). We used Tobii Pro Lab software with a Tobii Pro Nano eye tracker (60 Hz) to record eye movements.

VS-RTA was conducted immediately after gameplay. Participant was shown the gameplay recording and prompted to recall his/her thoughts, decisions, and rationale for their actions. Upon completion of the VS-RTA interview, the same participant participated in the ET-RTA interview. In ET-RTA, the participant viewed the same recording but overlaid with recorded eye-gaze data (fixation points, saccades, and gaze paths). These interviews were recorded using OBS screen recorder. Figure 1 shows the VS-RTA in the left image and the ER-RTA in the right image. Participant provided informed consent for video recording, eye-tracking data collection, and use of anonymized data for research purposes.

All verbalizations were transcribed and analyzed by the first author, focusing on the interviewer’s questions and the participant’s responses. Using content analysis, we mapped the transcribed verbalizations to the corresponding gameplay to identify comparable scenes from both interview conditions’ thematic analysis. Resulting themes were discussed with Authors 2 and 3 to reduce bias and enhance the validity of findings.

3. Result

Our analysis compared verbal responses from VS-RTA and ET-RTA conditions across equivalent gameplay segments, revealing systematic differences in response quality, quantity, and cognitive depth. Quantitative analysis revealed significant differences in verbal productivity. During VS-RTA, the participant concluded the interview early, but continued elaborating when eye gaze patterns were shown. ET-RTA session yielded 1,233 words compared to 562 in VS-RTA, which amounts to more than double the verbal output. While no statistical inference is possible with a single case, this descriptive difference suggests a potential trend worth further investigation. The systematic comparison of equivalent gameplay segments identified four primary themes where ET-RTA demonstrated advantages over VS-RTA.

3.1 Reduction of Speculative Verbalizations

Participants frequently used tentative language in the VS-RTA, demonstrating uncertainty about actual thoughts during gameplay. For instance (scene 1: Figure 1), when discussing diesel generator placement, the participant stated: *"Probably I was thinking that it's too near from the city. It will pollute. So I put it far. Maybe that is the reason I thought."* This excerpt illustrates the speculative nature of recall, with phrases like "probably" and "maybe". However, when eye-gaze patterns were shown for the identical instance, the same participant provided more definitive statements: *"I was looking for the position of the diesel generator that will be appropriate. I was checking the distance from the city...".* This shows that ET-RTA helps eliminate speculative language and replaces uncertainty with concrete descriptions. This demonstrates how eye-gaze patterns provided concrete anchors for more precise recall.

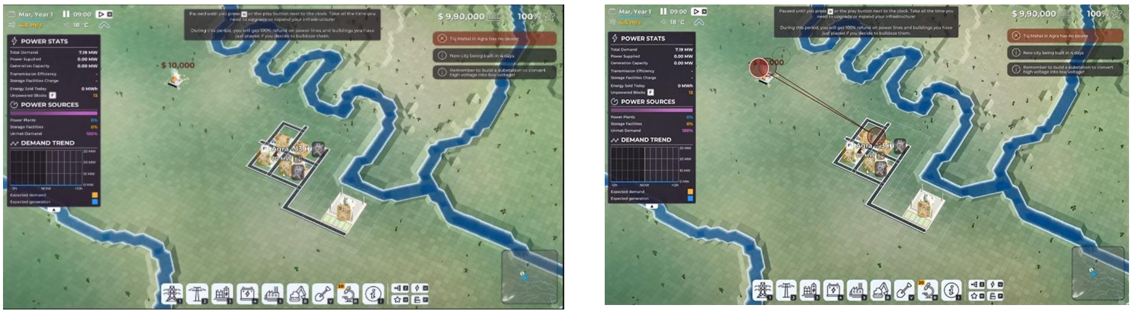

Figure 1. Scene 1 of gameplay for VS-RTA (left) and for ET-RTA (right) with fixations (red circles) and saccades (line connecting fixations)

3.2 Precision in Attentional Mapping

VS-RTA requires researchers to infer attentional focus from verbal descriptions and screen interactions, making this process vulnerable to confirmation bias. In this case, interview questions are primarily based on activities visible on-screen. In Scene 2, during VS-RTA, researchers formed questions based on observed screen selections: *"You selected the diesel generator? What made you choose it?"* To which the answer was: *"I was trying to put it near city. As its generating capacity was more..."*. In contrast, ET-RTA's eye-gaze data revealed the participant's visual attention patterns, enabling researchers to formulate targeted probes tailored to a specific fixation pattern (the participant's gaze hovered only on the selection button without reading the plant name and specifications). This observation led to a more precise, evidence-based question: *"While selecting this diesel generator, you haven't checked the full name and the ratings, you just selected it. What made you choose this option?".* This targeted probe elicited a markedly different and more honest response: *"I have not explored much, I saw a diesel generator. I just selected that..."*. This comparison demonstrates how ET-

RTA enables researchers to move beyond assumptions about participant attention and formulate questions that accurately reflect actual cognitive processes.

3.3 Mitigation of Reconstruction Artefacts

Participants in VS-RTA may describe what they are currently observing on screen, rather than recalling their actual thoughts. These reconstruction artefacts are particularly evident when describing static screens with minimal interaction. During Scene 3, when the participant began simulating grid operation, their verbalization focused exclusively on the visible changing numerical values: *"...transmission efficiency is not that good, the efficiency is 84-85%, very poor. Transmission efficiency should be greater than 95-96..."* This response demonstrates typical reconstruction behavior, describing observable screen elements without providing insight into the underlying decision-making process or attentional shifts. In contrast, ET-RTA with eye-gaze data showcases that the participant's attention pattern was more complex. The gaze data revealed the participant initially fixated on power statistics but then shifted attention to "Agra city." This evidence enabled rich verbalization of actual in-situ thought processes: *"...I was thinking the transmission efficiency should be high and capacity utilization should be high... because of Agra town is nearby, and it has higher capacity... Look this Taj Mahal is not powered. I was thinking to power up that. I was looking at power stats and the city because we have to give supply at higher efficiency.".* This response reveals the participant's actual strategic thinking for connecting transmission efficiency to geographic considerations and customer needs instead of simply describing visible numbers. This comparison illustrates how eye-gaze mitigates reconstruction artefacts by providing evidence of attention patterns, enabling participants to verbalize authentic cognitive processes rather than post-hoc rationalizations based on current screen observations.

3.4 Eliciting Granular Process-level Data

Participant often describe on-screen actions during VS-RTA, providing surface-level descriptions of what they did and basic reasoning. When confronted with eye gaze patterns, it provides a concrete visual representation of their cognitive processes. Reviewing eye movements enables a deeper understanding of attention patterns, decision-making, and problem-solving. In Scene 4, during VS-RTA, the participant provided a general description of the fault-recovery process: *"Probably something low voltage power line has failed as Taj Mahal, Agra has lost power, customer satisfaction has come down. I thought to provide on parallel, as I try to repair it...".* This response demonstrates a typical problem description with speculative language and limited insight into the detailed decision-making process. In contrast, when shown eye-gaze patterns revealing prolonged fixation on the "FAULT" indicator and revisits of repair costs, the participant's verbalization became significantly more granular and process-oriented: *"Here the line fault has occurred on the main line... I was thinking about the reasons of fault occurrence as I saw the customer satisfaction is going down. I repaired the fault for some $500. But it failed again, so again I was looking at how many funds are available... I thought to provide the second parallel transmission line, and share the load, it will reduce the heat, and fault won't reoccur...".* This reveals the complete cognitive sequence, i.e., fault identification, causal reasoning, cost consideration, outcome evaluation, and alternative solution generation. This demonstrates that eye gazes enabled the participant to articulate fine-grained problem-solving steps absent from the VS-RTA. This comparison illustrates how reviewing eye movements leads to increased self-awareness and improved verbalization of granular information about cognitive processes, transforming general action descriptions into detailed process-level insights.

4. Discussion and Conclusion

This exploratory case provides initial evidence that combining eye gaze information with RTA can enrich think-aloud protocols. While the findings are illustrative rather than generalizable, they point to promising directions for larger-scale validation. The four major themes that emerged in this study are enhanced memory precision through gaze cues, reduced attentional ambiguity, minimized cognitive reconstruction artifacts, and improved elicitation of process-level data. These align with established cognitive theories while offering novel protocol analysis for educational research. Previous research indicates that subjects develop fabricated thinking processes during stimulated recall due to incomplete memories and the human tendency to react rather than recall (Lyle, 2003; Güss et al., 2018). Our findings regarding minimized reconstruction highlight how eye gazes pinpoint attentional behavior for precise thought process verbalizations, providing crucial insights for designing effective learning environments and adaptive systems.

Flavell (1979), in his seminal work, highlighted that metacognitive experiences are frequently triggered when individuals are confronted with unfamiliar situations or challenging questions. Upon activation, these metacognitive experiences become a vital component of the cognitive monitoring system, providing crucial feedback and prompting strategic action (both cognitive and metacognitive). Our research indicates that displaying participants' eye-gaze data is **a potent metacognitive trigger** within the ET-RTA protocol.

VS-RTA faces significant challenges with non-veridicality, as granular details are not readily visible to interviewees or interviewers (Johnson et al., 2023; Ericsson & Simon, 1993). Our findings (in section 3.4) demonstrate that ET-RTA directly addresses this issue by externalizing implicit cognitive processes. Crucial cognitive actions that participants omitted were evident in eye-gaze patterns. Enhanced granularity corroborates Ericsson and Simon's (1993) assertion that verbal reports primarily reflect cognitive products unless augmented by externalized cues. Our findings extend this principle by demonstrating how gaze visualization targets the process dimension of cognition, making implicit evaluation criteria explicit and verifiable. This documentation is particularly critical in game-based learning contexts, where iterative strategy refinement constitutes a core learning mechanism.

Eye-gaze patterns trigger metacognitive activation, prompting deeper reflection on attentional processes and cognitive states. Participants observing eye movements on inactive screen elements recalled underlying monitoring behaviors, suggesting externalized gaze data acts as a catalyst for self-reflection and richer verbal data. Our quantitative and qualitative (section 3.4) analyses demonstrate ET-RTA's effectiveness in triggering metacognition and motivating richer verbal responses.

VS-RTA relies on participants' short-term memory during interviews, which organizes information in small chunks, limiting recall accuracy (Ericsson and Simon 1993; Johnson et al., 2023). Our study found VS-RTA's unassisted memory recall resulted in tentative language (section 3.1) as they struggled connecting video cues with fading memories. ET-RTA's gaze replays provided specific visual reference points that reduced tentative language and increased clarity, which helps overcome the limitation of VS-RTA regarding the temporal decay of episodic memory.

5. Conclusion and Limitation

Through a detailed case, this study illustrates how integrating eye-gaze data into retrospective think-aloud may help address some limitations of video-stimulated protocols. Findings overcome the challenges of VS-RTA, such as memory decay and cognitive reconstruction. Two of our major themes highlight the methodological contributions of this approach: first, it demonstrates that eye-gaze can act as a potent **metacognitive trigger**, leading to the elicitation of granular cognitive process data; and second, it streamlines the interview process by enabling the formulation of targeted probes based on specific fixation patterns. These findings have significant pedagogical implications, as ET-RTA can be adapted as a learning tool to help students develop better self-regulation and strategic thinking skills. By extending the application of ET-RTA beyond its traditional use in usability studies, our research provides new insights into its potential for educational contexts.

The within-subjects, single-participant design constrains generalizability, and the fixed order of protocols introduces possible practice and order effects. As such, the results should be viewed as exploratory. Nevertheless, the analysis demonstrates the feasibility of ET-RTA in educational contexts and generates hypotheses for larger, counterbalanced studies. Future research should explore a broader range of learning contexts and a larger sample size to provide a more comprehensive list of ET-RTA's benefits in learning and instruction.

Acknowledgements

This work is supported by the TIH Foundation for IoT and IoE under grant ID RD/0123-TIH0018-003.

We thank Federico Casares, Founder of Rhombico Games, for allowing us to use “The Power to the People” game for our study.

References

Alhadreti, O., & Mayhew, P. (2018, April). Rethinking thinking aloud: A comparison of three think-aloud protocols. In Proceedings of the 2018 CHI conference on human factors in computing systems (pp. 1-12).

Dewi, G. A. P. A. (2023). *USABILITY TESTING PADA APLIKASI PERSEDIAAN KABUPATEN BADUNG MENGGUNAKAN TEKNIK PERFORMANCE MEASUREMENT, RTA, DAN SUS* (Doctoral dissertation, Universitas Pendidikan Ganesha).

Elbabour, F., Alhadreti, O., & Mayhew, P. (2017). Eye tracking in retrospective think-aloud usability testing: Is there added value?. *Journal of Usability Studies*, *12*(3), 95-110.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological review*, *87*(3), 215.

Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis (revised edition). Overview of methodology of protocol analysis.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist, 34, 906 - 911.

Güss, C. D., Tuason, M. T., Göltenboth, N., and Mironova, A. (2018). Creativity through the eyes of professional artists in Cuba, Germany, and Russia. J. Cross Cult. Psychol. 49, 261–289. doi: 10.1177/00220221177308172.

Johnson, W. R., Artino Jr, A. R., & Durning, S. J. (2023). Using the think-aloud protocol in health professions education: an interview method for exploring thought processes: AMEE Guide No. 151. *Medical teacher*, *45*(9), 937-948.

Lyle, J. (2003). Stimulated recall: A report on its use in naturalistic research. British educational research journal, 29(6), 861-878.

Noushad, B., Van Gerven, P. W. M., & de Bruin, A. B. H. (2023). Twelve tips for applying the think-aloud method to capture cognitive processes. *Medical Teacher*, *46*(7), 892–897. https://doi.org/10.1080/0142159X.2023.2289847

Olsen, A., Smolentzov, L., & Strandvall, T. (2010, September). Comparing different eye tracking cues when using the retrospective think aloud method in usability testing. In Proceedings of HCI 2010. BCS Learning & Development.

Paikrao, A. M., & Mitra, R. (2023). Expert-Novice Differences In Electrical Circuit Analysis Based On The Order Of Attention. European Society for Engineering Education (SEFI). DOI: 10.21427/G9YB-5J15

Pathan, R., Murthy, S., & Rajendran, R. (2025). Identifying cognitive and metacognitive processes using retrospective and concurrent think-aloud protocols. In B. Jiang et al. (Eds.), Proceedings of the 33rd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education.

Rajendran, R., Kumar, A., Carter, K. E., Levin, D. T., & Biswas, G. (2018). Predicting Learning by Analyzing Eye-Gaze Data of Reading Behavior. *International Educational Data Mining Society*.

W. Rainey Johnson, Anthony R. Artino Jr & Steven J. Durning (2023), Using the think aloud protocol in health professions education: an interview method for exploring thought processes: AMEE Guide No. 151, Medical Teacher, 45:9, 937-948, DOI: 10.1080/0142159X.2022.2155123