

Exploring Digital Learning Processes through Webcam-Based Eye Tracking: Development and Applications of SNUWET

Siwon Sung, University of California, Irvine, siwons1@uci.edu
Hyerin Ryu, University of Maryland, hyerin@umd.edu
Juno Hwang, Seoul National University, wnsdh10@snu.ac.kr
In Chull Jang, Seoul National University, icjang@snu.ac.kr

Abstract: This paper introduces webcam-based eye tracking as a methodological alternative for exploring the multimodal nature of digital learning processes and presents Seoul National University Webcam-based Eye Tracking (SNUWET), a browser-based qualitative research suite. Two illustrative cases, one empirical study of multimodal meaning-making and one pedagogical application in pre-service teacher education, demonstrate webcam-based eye tracking's potential for moment-to-moment process tracing in authentic learning contexts, along with its methodological implications.

Introduction

Digital platforms now generate extensive traces enabling large-scale analyses of learner engagement. While such datasets provide valuable insights into online behavior, they reveal little about how learning unfolds moment by moment as learners navigate and make decisions within complex digital environments (Baker et al., 2020; He & Cui, 2025). Capturing these processual dimensions requires methodological tools that can observe learning as it happens, with minimal intrusion. While process-tracing technology such as screen recording and keystroke logging reveal procedural aspects of learning (Séror, 2013), they primarily capture overt manual actions leaving opaque gaze behaviors that precede and guide those actions. Without gaze-level data, process-tracing research remains limited in gaining holistic understanding of learners' multimodal engagement with the diverse resources that mediate learning in digital environments (Sung & Jang, 2024).

Eye tracking has long been used in research across language processing (Godfroid, 2019; Michel et al., 2020), human–technology interaction (Korkmaz, 2026), and technology-mediated learning (Cappellini & Hsu, 2022). Despite its value, conventional laboratory-based systems are costly and limited to artificial settings, thereby constraining ecological validity and accessibility. Addressing these limitations, webcam-based eye tracking utilizes built-in webcams to unobtrusively capture gaze data across authentic learning environments—classrooms, homes, and informal spaces (Yang & Krajbich, 2021). This paper introduces webcam-based eye tracking as a methodological alternative for qualitatively exploring the micro-processes of digital learning. It presents SNUWET, a browser-based qualitative research suite that makes process tracing accessible to educators and researchers, and reports one empirical and one pedagogical use case illustrating its potential as an educational research suite.

SNUWET: A webcam-based eye tracking tool for educational research

Developed by a research team in the Department of English Language Education at Seoul National University, SNUWET is a browser-based qualitative research suite for collecting multimodal data, including screen and audio recordings and eye-tracking data, across formal and informal learning environments. Grounded in a design-based implementation research (DBIR) approach, it examines the usability, validity, and educational affordances of webcam-based eye tracking in authentic contexts. Through iterative design and research cycles across K–12, graduate, and teacher-education settings, the system was refined to ensure usability for both researchers and participants.

The interface features a two-stage pipeline consisting of a recording platform and a post-analysis tool. The recording webpage (snuwet.io) allows users to capture webcam, screen footage, and microphone audio, which are saved locally or to cloud storage. MediaPipe (Grishchenko et al., 2020) is used for face-landmark detection to monitor user positioning and gaze stability, generating multimodal datasets comprising user metadata, video, and audio. For post-analysis, a Python-based tool extracts and calibrates gaze data and generates gaze-plot visualizations. L2CS-Net (Abdelrahman et al., 2023) is employed for yaw and pitch estimation, addressing the limitations of standard facial-landmark models in gaze estimation. These estimates are fitted to calibration videos using a regression model that accounts for individual webcam positioning and screen size, producing corrected gaze coordinates. The post-analysis tool operates in Google Colab, reflecting iterative design refinements from an initial standalone Python script to address operating-system constraints and support accessible, cloud-based use. The resulting interface enables researchers and educators to conduct

learning analyses through an accessible, browser-based workflow, reflecting SNUWET's broader commitment to expanding methodological access and participation in webcam-based eye tracking research.

Applying webcam-based eye tracking to learning and teaching contexts

Case 1: Researching multimodal meaning-making in digital text comprehension

Forty-four eighth-grade EFL (English as Foreign Language) students viewed three digital advertisements in a school computer lab while gaze data were collected via SNUWET, after which they responded to three comprehension questions designed to assess their interpretation of advertising intent. The 10 participants whose recordings yielded the highest gaze accuracy were subsequently selected for stimulated-recall interviews, in which they verbalized their thought processes while watching their own gaze-plot recordings. Drawing on multimodal discourse analysis, participants' reading paths (Kress & Van Leeuwen, 2006) were examined through an iterative qualitative coding process applied to both data sources. One learner (Cecil) actively attended to slogans, facial images, and brand logos yet produced a literal rather than inferential reading of the advertisement (see Figure 1), whereas other learners focused primarily on visual elements with less attention to linguistic slogans, yet correctly inferred the intended meaning. These contrasting patterns demonstrate that visual attention to multimodal elements does not automatically translate into successful comprehension, but meaning-making involves active integration across semiotic modes. Crucially, these processes were only made visible through the real-time screen recording triangulated with gaze data and stimulated recall, illustrating how the method can be employed in classroom contexts to collect ecologically valid data on learning processes as they become increasingly mediated by digital technology.

Case 2: Pedagogical use in pre-service teacher education

As part of a methods course at an English teacher education program in South Korea, 14 pre-service teachers participated in a two-hour learning analytics workshop using SNUWET. They were tasked with conducting a learning analytics where they recorded 20 minutes of online English learning activities of EFL learners, conducted stimulated recall interviews, and wrote an analytical report tracing how learners attended to multimodal cues and employed strategies in a digital learning environment, along with discussions of pedagogical implications. In one instance, a pre-service teacher identified a "text first, image second" strategy in reading an English webtoon, where the learner reread surrounding text for vocabulary inference before consulting images (see Figure 2). Suggesting the implications of guided vocabulary teaching using webtoon, the pre-service teacher acknowledged the analytical usefulness of the gaze data. By making visible the hidden processes of language learners' interpretation of multimodal text, and with its accessibility and usability, SNUWET enables teachers to engage in evidence-based pedagogical reasoning about how to support language learners in digital environments.

Figure 1

Screen Capture of Cecil's Gaze Plot Recording

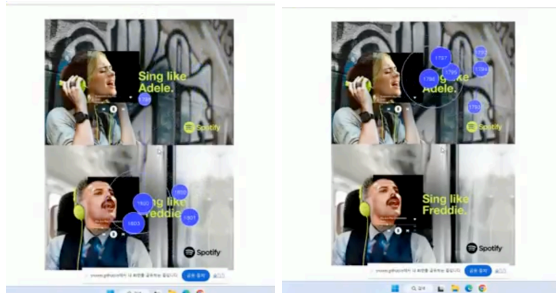


Figure 2

A Pre-service Teacher's Learning Analytics Report

1) 텍스트 중심, 그림 보조 전략: 대부분의 어휘 수준에서 참여자는 우선적으로 해당 어휘가 포함된 문장 및 앞뒤 텍스트 문맥을 반복적으로 읽으며 의미를 파악하려 했다. 이후, 그림 정보는 자신이 텍스트를 통해 추론한 의미를 확인하거나, 텍스트만으로 의미 파악이 어려울 때 추가적인 단서를 얻기 위한 보조적 수단으로 활용되었다.

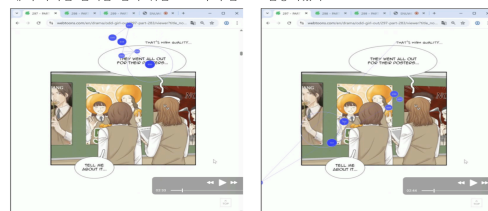


그림 1 'went all out'의 의미 추론 시점 - 텍스트에 우선 집중한 후 그림에 집중

Conclusion

With two empirical illustrations, this study presents SNUWET as both a research and pedagogical tool that enables ecologically valid qualitative exploration of learning processes in digital contexts. By incorporating gaze data, SNUWET enables a more holistic understanding of learners' moment-to-moment digital learning processes. Moreover, its accessibility and ease of use open new avenues for creative collaborations among researchers and practitioners. Nevertheless, it is worth noting that there is a trade-off between accuracy and ecological validity, suggesting that users strengthen interpretations through triangulation with complementary data sources, such as stimulated recall interviews, to achieve rigorous and contextually grounded research.

References

- Abdelrahman, A. A., Hempel, T., Khalifa, A., & Al-Hamadi, A. (2023). L2CS-Net: Fine-grained gaze estimation in unconstrained environments. *2023 8th International Conference on Frontiers of Signal Processing (ICFSP)*, 98–102.
- Baker, R., Xu, D., Park, J., Yu, R., Li, Q., Cung, B., Fischer, C., Rodriguez, F., Warschauer, M., & Smyth, P. (2020). The benefits and caveats of using clickstream data to understand student self-regulatory behaviors: Opening the black box of learning processes. *International Journal of Educational Technology in Higher Education*, 17(1), 13.
- Cappellini, M., & Hsu, Y.-Y. (2022). Multimodality in webconference-based language tutoring: An ecological approach integrating eye tracking. *ReCALL*, 34(3), 255–273.
- Godfroid, A. (2019). *Eye Tracking in Second Language Acquisition and Bilingualism: A Research Synthesis and Methodological Guide*. Routledge.
- Grishchenko, I., Ablavatski, A., Kartynnik, Y., Raveendran, K., & Grundmann, M. (2020). Attention Mesh: High-fidelity face mesh prediction in real time. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*.
- He, S., & Cui, Y. (2025). A systematic review of the use of log-based process data in computer-based assessments. *Computers & Education*, 228, 105245.
- Korkmaz, A. (2026). Mapping Eye-Tracking Research in Human–Computer Interaction: A Science-Mapping and Content-Analysis Study. *Journal of Eye Movement Research*, 19(1), 23.
- Kress, G., & Van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
- Michel, M., Révész, A., Lu, X., Kourtali, N.-E., Lee, M., & Borges, L. (2020). Investigating L2 writing processes across independent and integrated tasks: A mixed-methods study. *Second Language Research*, 36(3), 307–334.
- Séror, J. (2013). Screen capture technology: A digital window into students' writing processes/Technologie de capture d'écran: Une fenêtre numérique sur le processus d'écriture des étudiants. *Canadian Journal of Learning and Technology/La Revue Canadienne de l'Apprentissage et de la Technologie*, 39(3), 1–16.
- Sung, S., & Jang, I. C. (2024). South Korean STEM graduate students' use of ChatGPT in self-initiated L2 writing: A process-tracing Study. *Korea Journal of English Language and Linguistics*, 24, 1415–1435.
- Yang, X., & Krajbich, I. (2021). Webcam-based online eye-tracking for behavioral research. *Judgment and Decision Making*, 16(6), 1485–1505.

Acknowledgements

This work was funded by the National Research Foundation of Korea (NRF-2024S1A5A8026540).