Examining Students' Intrinsic Cognitive Load During Program Comprehension – An Eye Tracking Approach

Magdalena Andrzejewska^(✉) and Agnieszka Skawińska

Pedagogical University, Cracow, Poland
magdalena.andrzejewska@up.krakow.pl

Abstract. Programming as a cognitive activity requires the utilization of various kinds of mental models that involve different cognitive loads while students learn to program. The article discusses the results of an experiment aimed at answering the following question: are eye tracking based measures related to the intrinsic cognitive load (ICL) connected with program comprehension? Thirty one students of computer science took part in the experiment. They analyzed two program codes written in the C++ language to search for (1) logical errors (LER) and (2) syntax errors (SER). ICL was measured by subjective rating of the difficulty of each task. There were significant differences found for the subjective measures of intrinsic load, the effectiveness and the time of tasks performance, and the values of eye tracking parameters: fixation duration average (FDA) and saccade amplitude average (SAA) in two experiment conditions. Longer fixation and shorter saccades were associated with higher ICL. The results obtained suggest that FDA and SAA are eye tracking measures sensitive of intrinsic cognitive load.

Keywords: Cognitive load · Program comprehension · Eye tracking

1 Introduction

Studies conducted for many years consistently show that acquiring the skill of programming at its early stage poses difficulty to students [incl. 9]. Programming is a complex skill that, on the one hand, encompasses mechanisms of problem-solving and algorithm construction, and, on the other, demands knowledge of the syntax and semantics of the programming language [14]. It is, therefore, assumed that the experienced difficulties result to a significant extent from the excessive cognitive load (CL) occurring in the process of learning [15].

This paper approaches cognitive load as a triarchic concept, such as the one defined within the framework of the Cognitive Load Theory (CLT) which distinguish three types of cognitive load: intrinsic load (ICL) (related to the difficulty of a task, its structure or complexity, referring to an individual's effort load needed to learn a concept), extraneous load (ECL) (related to information presentation and instructional format), and germane load (GCL) (referring to the mental resources involved in

acquiring and automating schemata in the long-term memory [16]. Designing education-related materials following the principles of CLT and measuring cognitive load values has seen a growing interest in the field of research in recent years. A few of these studies have attempted to examine the application of the cognitive load theory in computer science education – especially in teaching programming [1, 11, 12, 17]. But despite many conducted studies, the problem of how to measure the cognitive load occurring during learning is still widely discussed [13]. Researchers are looking for measures designed to distinguish between the different types of load (ICL, ECL, GCL) [incl. 7, 10].

There are four dominant types of methods to address the measurement of cognitive load: subjective rating, performance-based measures, physiological measures, and behavioral measures [3]. Among the physiological measurements of CL, eye-based measures appear to be the most popular. The most common eye tracking measures of CL there are: changes in pupil size, blink rate and duration, saccade speed, and fixation duration [incl. 4, 8]. But it should be also mentioned that there are no threshold values of these indices that would allow for making inferences regarding the actual level of CL. Eye tracking methods have been shown to distinguish between tasks involving low cognitive loads and tasks involving high cognitive loads [5]. It has also been examined how cognitive load factors can be independently measured with eye tracking methods as well as how they are related to the subjective rating scale [6, 18]. But there seems to be an underrepresentation of eye tracking research that would apply to programming tasks, and – in particular – research investigating which eye movement parameters are sensitive to different types of cognitive load in the process of learning to program.

2 Current Study

The studies conducted so far have not yet analyzed the cognitive load involved in programming activities such as code debugging, in conditions where (1) the study subjects analyze a code without using an Integrated Development Environment (IDE) (where they can trace and run the program, which leads to the occurrence of additional factors disturbing the comprehension of the program) and (2) they analyze the exact same code but perform two different cognitive tasks – such as (a) searching for logical errors (LER) and (b) searching for syntax errors (SER). Given the above and with respect to the CLT principles, an assumption can be made that extraneous load (ECL)—which is related to the instructional format—should not differ between the two task versions. Therefore, this experiment design will be mainly related to ICL which is affected by the level of difficulty of the concept related to its complexity. It is considered that the subject's prior knowledge determine the ICL [10]. These assumptions are similar to those adopted by [2].

In the light of the above, our research question is: what eye movement parameters are sensitive to intrinsic cognitive load that program comprehension imposes on a student? To address the research question, we examined several fixation and saccade parameters, excluding those that were correlated with each other. Finally, we focused on fixation duration and saccade length that were assumed to be the measures of the total cognitive load [8], i.e.: fixation duration average (ms) (FDA: the sum of the

duration of all fixations divided by the number of fixations) and saccade amplitude average ($^{\circ}$) (SAA: the sum of all saccade amplitudes divided by the number of saccades in the trial). Our analysis also included: time (ms), which refers to the number of milliseconds spent answering each task, and accuracy (%), meaning the percentage of errors reported by the subjects. These variables are also included in the research as performance-based measures of ICL [incl. 2].

3 Method

Experimental apparatus. Our study was conducted using the iViewX Hi-Speed eye tracker manufactured by SensoMotoric Instrument (SMI). The following SMI software was used to prepare the experiment and compile its results: Experiment Center and BeGazeTM 2.4.

Participants. Thirty four students of computer science participated in the study. The results of 3 subjects were removed from our analyses due to eye tracking measurement errors. The final sample resulted in 31 participants and consisted of 23 men and 8 women, aged between 21 and 29 ($M = 23.90$, $SD = 1.66$). All students completed a C++ programming course and had previously learned the concepts that were employed in the tasks they were asked to perform.

Procedure and material. After the subjects were familiarized with the experimental procedure, the eye tracking system was calibrated and validated. Next, each participant received two codes of short but complete programs written in C++. Each program offered a solution to the same problem, which was the implementation of an algorithm of sorting a ten-element table based on the *selection sort* method in a non-decreasing order. There were two separate programs that were presented in the same sequence to each participant. The first program contained four only logical errors (LER), the second code contained five only syntax errors (SER). Students were asked to find errors in both coding tasks and provide an answer orally. The codes were neither compiled nor run. The subjects had unlimited time to find the errors. In a short post-survey, study participants rated the difficulty level related to each task and their programming skills level (on a Likert scale from 1 (very easy/low) to 5 (very difficult/high)).

4 Results

Most of the students considered their programming skills to be on a medium level ($M = 2.80$, $SD = 0.7$, $Me = 3$, $Q1 = 2$, $Q3 = 3$); the sample seems to be quite homogeneous with respect to this feature. In the case of subjective measurement, the LER task imposed a higher intrinsic load as compared to the SER task (see: Table 1). Students rated searching for logical errors as more difficult than searching for syntax errors.

We studied the distribution of the gaze data: FDA, SAA, the performance data: Time, Accuracy, and Difficulty rating using the Shapiro-Wilk test, and found that only the FDA parameter followed the theoretical normal distribution (LER: $W = 0.943$, $p = 0.102$; SER: $W = 0.942$, $p = 0.097$). Thus, we decided to use a paired t-test for FDA and the Wilcoxon signed-rank test as a non-parametric test for the remaining variables.

Table 1. Wilcoxon test and paired t-test for the dependent variables

Variable	M			
	LER (high ICL)	SER (low ICL)	Z	p
Difficulty rating	3.6	2.8	3.587	0.000
Accuracy (%)	23.4	37.4	3.250	0.001
Time (ms)	238091.6	146887.5	3.155	0.002
SAA (°)	4.5	4.7	2.027	0.043
			t	p
FDA (ms)	251.5	233.6	4.602	0.000

If we refer to Table 1, we can see that there are significant differences both in the time and the task performance during searching for syntax versus logical errors. In the case of LER (high ICL), the subjects spent more time and found fewer errors compared to SER (low ICL). Furthermore, we found that students had a significantly higher FDA and a significantly lower SAA in the LER task (high ICL) compared to the SER task (low ICL), which suggests that these eye-based parameters are sensitive to ICL.

5 Conclusions

The outcomes of our study show that (1) FDA and SAA differed significantly in two task conditions, and that (2) longer fixation and shorter saccades were associated with a higher intrinsic cognitive load. The obtained findings suggest that these eye tracking measures are sensitive to ICL and therefore are a promising indicator of ICL related to the specific mental process of program analysis aimed at identifying logical and syntax errors. However, it was a preliminary study and therefore has some limitation that should be taken into consideration and addressed in future works. The aspects that need to be taken into account include: (1) increasing the number of subjects and comparing novice and expert results; (2) extending the scale of the subjective load assessment; (3) entering code difficulty levels; (4) introducing redundancy to measure ECL; (5) examining how ICL and ECL change in time intervals.

References

1. Asai, S., Phuong, D.T.D., Harada, F., Shimakawa, H.: Predicting cognitive load in acquisition of programming abilities. *Int. J. Electr. Comput. Eng.* **9**(4), 3262–3271 (2019)
2. Ayres, P.: Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learn. Instr.* **16**(5), 389–400 (2006)
3. Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S.Z., Khawaji, A., Conway, D.: Robust Multimodal Cognitive Load Measurement. Springer, Cham (2016). <https://doi.org/10.1007/978-3-319-31700-7>
4. Chen, S., Epps J., Ruiz, N., Chen, F.: Eye activity as a measure of human mental effort in HCI. In: Proceedings of the 16th International Conference on Intelligent User Interfaces, pp. 315–318. Association for Computing Machinery, Palo Alto (2011)
5. Chen, S., Epps, J.: Using task-induced pupil diameter and blink rate to infer cognitive load. *Hum.-Comput. Interact.* **29**(4), 390–413 (2014)
6. Debye, N., Leemput, C.: What does germane load mean? An empirical contribution to the cognitive load theory. *Front. Psychol.* **5**, 1099 (2014)
7. DeLeeuw, K.E., Mayer, R.E.: A comparison of three measures of cognitive load: evidence for separable measures of intrinsic, extraneous, and germane load. *J. Educ. Psychol.* **100**(1), 223–234 (2008)
8. Holmqvist, K., Nyström, N., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J.: Eye tracking: a comprehensive guide to methods and measures. Oxford University Press, Oxford (2011)
9. Konecki, M.: Problems in programming education and means of their improvement. In: Katalinic, B. (ed.) DAAAM International Scientific Book 2014, pp. 459–470. DAAAM International, Vienna (2014)
10. Leppink, J., Paas, F., Van der Vleuten, C.P.M., Van Gog, T., Van Merriënboer, J.J.G.: Development of an instrument for measuring different types of cognitive load. *Behav. Res. Methods* **45**(4), 1058–1072 (2013)
11. Morrison, B.B., Dorn, B., Guzdial, M.: Measuring cognitive load in introductory CS: adaptation of an instrument. In: Proceedings of the ICER 2014: International Computing Education Research Conference, Glasgow, Scotland, UK, pp. 131–138 (2014)
12. Nolan, K., Mooney, A., Bergin, S.: Examining the role of cognitive load when learning to program. In: 3rd International Workshop on Eye-Movements in Programming, Joensuu, Finland, pp. 23–24 (2015)
13. Orru, G., Longo, L.: The evolution of cognitive load theory and the measurement of its intrinsic, extraneous and germane loads: a review. In: Longo, L., Leva, M.C. (eds.) H-WORKLOAD 2018. CCIS, vol. 1012, pp. 23–48. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-14273-5_3
14. Robins, A., Rountree, J., Rountree, N.: Learning and teaching programming: a review and discussion. *Comput. Sci. Educ.* **13**, 137–172 (2003)
15. Shaffer, D., Doube, W., Tuovinen, J.: Applying cognitive load theory to computer science education. In: Petre, M., Budgen, D., (eds.) 15th Annual Workshop of the Psychology of Programming Interest Group, pp. 333–346 (2003)
16. Sweller, J., Van Merriënboer, J.J.G., Paas, F.G.W.C.: Cognitive architecture and instructional design. *Educ. Psychol. Rev.* **10**(3), 251–296 (1998)

17. Yousoof, M., Sapiyan, M.: Cognitive load measurement in learning programming using NASA TLX rating scale (Non Physiological Measures). In: Proceedings of the 13th International Conference on Applied Computer and Applied Computational Science, Kuala Lumpur, Malaysia, pp. 235–245 (2014)
18. Zu, T., Hutson, J., Loschky, L.C., Rebello, N.S.: Use of eye-tracking technology to investigate cognitive load theory. In: Proceedings of 2017 Physics Education Research Conference, pp. 472–475. American Association of Physics Teachers, Cincinnati, OH (2018)