RT&A, Special Issue No. 7 (83),
Volume 20, May 2025

INVESTIGATING THE APPLICATIONS AND IMPACTS OF LARGE LANGUAGE MODELS IN CHEMISTRY EDUCATION

Canan Kocak Altundag[1], Sencer Yucel[1], Faxraddin Yusubov[2]

•

[1]Turkey, Ankara, Hacettepe University, Beytepe
[2]Azerbaijan Technical University, Baku, H. Cavid avenue, 25
canan.kocak@hacettepe.edu.tr, senceryucel@hacettepe.edu.tr, yusubov.faxraddin@aztu.edu.az

Abstract

Organic chemistry, which is an experimental science, deals with the acquisition and characterization of pure organic compounds from natural or synthetic pathways by appropriate methods. In addition to a good textbook in chemistry education, the laboratory of this course is also of great importance. For the realization of the processes, equipment made of various materials and their efficient use are required. This study has been prepared in order to make this laboratory more understandable and more enjoyable for students to recognize organic synthesis and see its products. In this article, Iodoform Synthesis experiment is taken as the basic experiment to explain the system. A virtual experimental environment powered by a Large Language Model (LLM) was created. The system presents students with a multi-step experiment which they are asked to complete correctly. The innovative side of this environment is how it interacts with the student when they make a mistake. This study was also conducted in an Organic Laboratory class, and thus created a research question: How did the idea of using Large Language Model in organic chemistry course affect student's mental workload?

Keywords: Laboratory instruction, large language model, chemistry education.

I. Introduction

Organic chemistry is a sub discipline of chemistry that investigates the synthesis, reactions and properties of compounds containing carbon in their structure Organic chemistry is basically an experimental field, the final decisions on the identity and structure of reaction products, their thermodynamic properties, and spectroscopic appearance are based on observation and measurement [1]. The laboratory is the place where students learn and apply the synthesis, separation, purification and analysis techniques of organic compounds and thus become acquainted with the experimental aspect of organic chemistry. However, the laboratory environment can be very complex for students. Students may have difficulty in performing experiments in organic chemistry laboratories. This study investigates the role of virtual experiment environment strengthened by an LLM in chemistry education. Artificial intelligence (AI) tools are spurring rapid developments in the chemical industry and academia [2]. Traditional learning methods in chemistry education, often limited to static textbooks and one-way knowledge transfer, have been changing lately to interactive AI tools.

II. The applications in organic chemistry laboratory

Students often use worksheets while conducting experiments in organic chemistry laboratory applications. With worksheets, known laws, principles, etc. can be revealed (induction) or verified (deduction) [3]. In the worksheet, the diagram of the experimental setup, the steps of the experiment, the collection of data and the evaluation of the data are given. Sometimes students can be expected to find the result by asking partially guiding questions [4]. These students who completed the experiment in accordance with the worksheet were not mentally challenged and did not succeed in becoming active participants of the experiment [5]. Because it is seen that a printed scenario can neither present specific laboratory techniques ready for reproduction nor prepare students for laboratory work well enough [6].

In order to provide students with a successful laboratory experience, it is necessary to use authentic organic chemistry applications that reflect organic chemistry practices and include modern techniques [7]. It is thought that especially prospective chemistry teachers who are not successful enough in experiments will react positively to such an innovative application in organic chemistry laboratory. Because learning can be more meaningful when the cognitive load caused by experimental work can be reduced [8].

One of the unique problems of students during organic laboratory applications is mental workload. Mental workload is defined as the amount of mental work required for a person to complete a task over a certain period of time. The fact that organic chemistry laboratory courses are conducted via experiments creates a mental workload rather than a physical workload. Issues such as the temporal pressures on students caused by long lectures at the laboratory and experimental reports that must be completed by the due date, and coordinating with the courses by showing intense effort create mental workload on students.

III. The Large Language Models in Organic Chemistry Laboratory Development

Organic Chemistry Laboratory I and II courses include a multifaceted education in which the basic concepts of organic chemistry and the necessary scientific skills are acquired. For this study, a different laboratory day was prepared for 16 students. The students were not given laboratory worksheets to perform the iodoform synthesis experiment. The students were given only materials required to synthesis the iodoform. A special application was prepared for students that they could follow on their mobile phones. We created a virtual experiment environment strengthened by an LLM. As the LLM, OpenAI's ChatGPT-3.5-Turbo was used. The system presents students with a multi-step experiment which they are asked to complete correctly.

The Iodoform Synthesis experiment is taken as the base experiment to explain the system. The student is expected to complete the experiment in the correct order among the mixed steps. There are some locked steps as can be seen (Figure 1). Even in this relatively short experiment consisting of only 8 steps, a mistake in one critical step could lead to many different outcomes and confuse the student. Thus, we created locked steps and did not allow the students to make mistakes when they came to the locked step and forced them to find the correct step without prompting any question to the LLM. Locked steps change their state once they have been selected in the correct order.

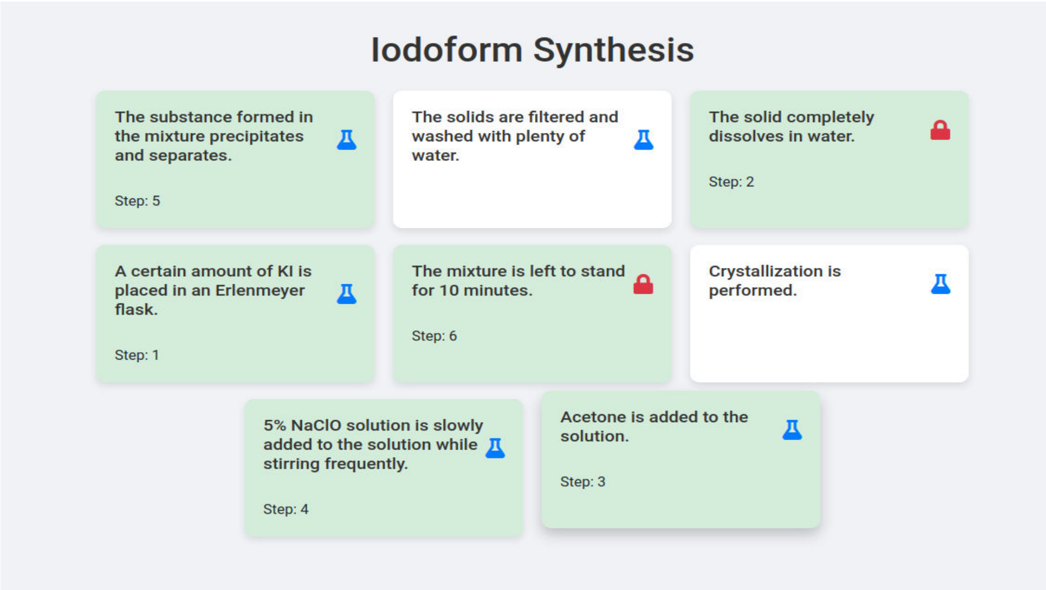

Figure 1: *The user interface seen by a student who has successfully reached step 7*

Below is an example of how the system feedback the student once they selected the step "The mixture is left to stand for 10 minutes." instead of "The substance formed in the mixture precipitates and separates.":

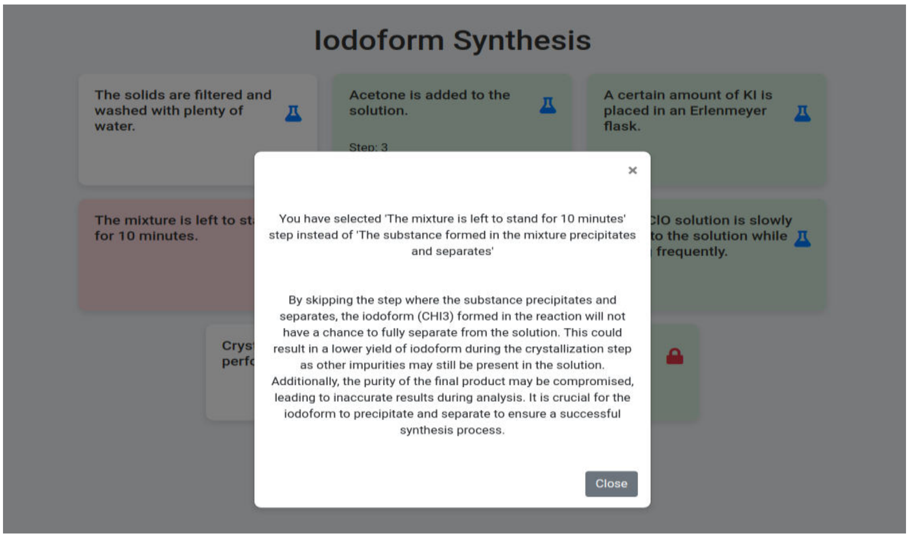

Figure 2: *LLM's response to a student who made a mistake in step 5 of the experiment*

Thanks to the working principle of LLMs, the system does not give the same answer when the same mistake is made again. It explains the same result with different words and expressions. This is a serious advantage as a student is likely to grasp something that he/she could not fully grasp before, using different expressions. Furthermore, there is no chance for the system to remain silent against a mistake.

IV. Data Collection Tool

This study, conducted in an Organic Laboratory Course through Large Language Model, had a research question. How did the idea of using Large Language Model in organic chemistry course affect student's workload? Several types of uni- and multidimensional subjective scales exist.

However, results from various studies have shown that NASA-TLX [9] is superior to SWAT in terms of sensitivity, especially for low mental workloads [10, 11]. Therefore, it was decided to use the NASA-TLX method to determine the mental workload of students (jobs required intense mental demands, physical requirements, and time constraints imposed by long-term analyses, etc.).

NASA-TLX is a multidimensional scale for which the overall mental workload is a function of mental demand, physical demand, temporal demand, performance, effort, and frustration dimensions, with each of these dimensions on a continuum. In the method, the workload score is determined in stages: scoring and ranking. Students in the first stage; considering the work they did, scored the workload they felt for the 6 dimensions by valuing a scale between 0 (very low) and 100 (very high) (divided into 5-point intervals) [12]. The determined scores are considered at five levels: very low (0-20), low (21-40), medium (41-59), high (60-79), and very high (80-100) [13].

The workload score was created by multiplying the scores given to the six criteria by the percentages of repetitions of the choices prioritized in the pairwise comparison. The NASA-TLX and mobile application solutions are evaluated according to the levels determined in the NASA-TLX solution.

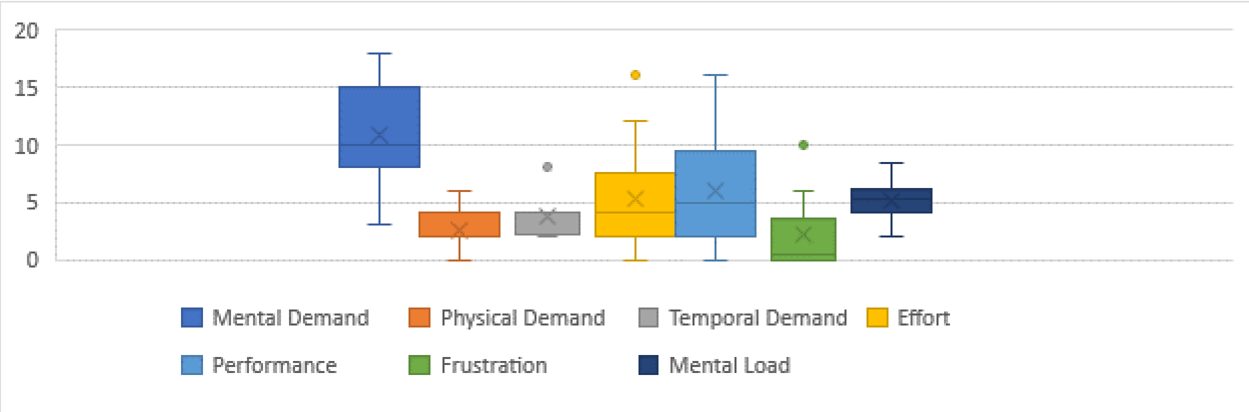

Figure 3: *NASA-TLX solution*

It was determined that the mental workload of 5 students was between 0-20 points, which is considered to be a very low level. It was determined that the mental workload of 11 students was between 21-40 points, which is considered to be at a low level.

IV. Student Views of the LLM

This study investigates student responses of virtual experiment environment strengthened by an LLM in an organic laboratory chemistry course. At the end of the study, the performance of application asked to the students. An 18-app survey questions, with scores ranging from 0 to 10, was administered to assess the students' experience. Student responses are shown in Table 1.

Table 1: *Students' responses by survey statement.*

Item		1	2	3	4	5	6	7	8	9	10
							%				
1	How do you rate your overall experience with our app?							18	37	12	33

2	How difficult is it to read the characters?	75	20		5						
3	How useful is our product for you?							12	18	33	37
4	How would you rate the user-friendliness of the product interface?							26	18		56
5	Considering that you use our interface extensively, how likely are you to recommend it to your friends and colleagues?							20	25	5	50
6	If you were to review the product, how would you rate it out of 10?							18	32	32	18
7	Overall, how easy did you find it to use?							8	10	32	50
8	How would you rate the usability of our app?								32	43	25
9	How would you rate your experience?						44			37	19
10	How would you rate the reliability of app?					1	18	18	10	12	37
11	How do you evaluate the quality of the content presented on the screen?					3	3	10	10	32	32
12	How would you rate the speed of our application?			3	3	10	10	32	32		
13	How would you evaluate the interface design of our application?					12	15	9	43	12	9
14	How useful are the displayed error messages?							13	13	13	61
15	How well does our app keep you informed about the progress of a task?								25	25	50
16	How difficult is it to read the characters on the screen?								19	19	62
17	How consistent is the use of terms across the application?							25	25	7	43
18	How much do you like the look of this app?					1	18	18	10	12	37

Student responses are shown in Table 1. Students engaged with LLM to follow experiment steps, revealing overall positive perceptions toward their usefulness. The responses in general were found to have high agreement in terms of playability, usability, and content and demonstrated that the LLM had positive features. Only item 12 showed that students were indecisive in speed of the application. This might stem from the fact that this was the first time these students were participating in a LLM about organic chemistry laboratory. In addition, they used the LLM with the slow internet.

VI. Results

Chemistry teacher candidates are given theoretical knowledge in organic chemistry I and II courses at the university, while experimental organic chemistry knowledge is explained comparatively in organic chemistry laboratory I and II courses. Students can learn both the application of theories and models in the interpretation and explanation of experimental results and the chemistry of organic compounds comparatively. Students derive their knowledge from the analysis of data collected by scientific methods. Therefore, laboratory applications are very important for chemistry education [14].

This study created a virtual experiment environment strengthened by an LLM. The LLM application includes the application steps of the 'Iodoform Experiment' that students should do in the organic chemistry laboratory. This application was a route that helped students to perform the steps of the experiment in order and was drawn to iodoform synthesis. The application aims to introduce students to modern chemistry through the use of artificial intelligence in organic chemistry laboratory. To date, despite the interesting progress made in artificial intelligence, artificial intelligence applications have not been common in chemistry teaching laboratories. The existence of an error rate should not be ignored in any AI system, but this is still not an obstacle to systems being able to use AI. With the development of artificial intelligence, it is inevitable that some of the traditional education methods we know will disappear, while others will change radically. In this first phase of LLMs, we see that learning and education can be significantly personalized. The use of LLMs in chemistry education offers a different learning style by providing a real-time feedback system in a virtual experimental environment.

As of the day this paper was written, one of the most advanced LLMs is ChatGPT from OpenAI. The most ideal way is to fine-tune an LLM for this specific purpose. Due to its requirements of serious resources, we did not go into fine-tuning, instead we built the system on ChatGPT in a way that can be considered a proof of concept. Overall, students in this study have a positive view of the utility of generative virtual experiment environment strengthened by an LLM in the organic chemistry laboratory course.

The minimum number of errors a student can possibly make in a virtual experiment consisting of 8 steps in total, 2 of which are locked, is 15. Similarly, for a 15-step experiment with 3 steps locked, the minimum number of errors that can be made is 66. It takes an unprecedented knowledge for a teacher to know what the consequences of all the different combinations of errors will be in all the different chemistry experiments. However, LLMs have already reached this level of knowledge, because the datasets they were trained on already contain the vast majority of errors that can be possibly made. One important application is the use of Large Language Models (LLMs) to provide instant feedback in virtual experimental environments. In this paper, we explored the role of LLMs in a virtual chemistry experimental environment and investigated their applications and impact on improving student learning outcomes. The innovative side of this environment is how it interacts with the student when they make a mistake. Rather than simply marking the answer as incorrect, the system stops the experiment and prompts an LLM to analyze the error and provide a technically detailed explanation. Research data were obtained from NASA-TLX. LLM application has enabled them to do the experiment in the remaining time of a course, and it has not burdened them with an intense study task and a very high mental workload. With LLMs, the effects of such a stressful workload could be alleviated and flexibility could be provided.

References

[1] Carey, F. A. & Sundberg, R. J. (2007). Advanced organic chemistry: part A: structure and mechanisms; Springer Science & Business Media.

[2] Baum, Z. J., Yu, X., Ayala, P. Y., Zhao, Y., Watkins, S. P. & Zhou, Q. (2021). Artificial Intelligence in Chemistry: Current Trends and Future Directions. J. Chem. Inf. Model, 61 (7), 3197–3212.

[3] Hand, B., & Treagust, D. F. (1991). Student achievement and science curriculum development using a constructive framework. School Science and Mathematics, 91(4), 172-176.

[4] Redfield, D. L., & Rousseau, E. W. (1981). A meta-analysis of experimental research on teacher questioning behaviour. Review of Educational Research, 51(2), 237-245.

[5] Sands, M. & Özçelik, D.A. (1997). Application studies in schools secondary education. Ankara: Agency.

[6] Pölloth, B., Teikmane, I., Schwarzer, S., & Zipse, H. (2019). Development of a modular online video library for the introductory organic chemistry laboratory. Journal of Chemical Education, 97(2), 338-343.

[7] Slade, M. C., Raker, J. R., Kobilka, B., & Pohl, N. L. (2014). A research module for the organic chemistry laboratory: Multistep synthesis of a fluorous dye molecule. Journal of Chemical Education, 91(1), 126-130.

[8] Bretz, S. L. (2001). Novak's theory of education: human constructivism and meaningful learning. Journal of Chemical Education, 78 (8), 1107.

[9] Hart, S. G., Wickens, C. D. (1990). Workload Assessment and Prediction. In MANPRINT: An Approach to Systems Integration. Dordrecht: Springer Netherlands, pp. 257-296.

[10] Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. Human Factors, 33,17- 33.

[11] Reid, G. B.; Nygren, T. E (1988). The Subjective Workload Assessment Technique: A Scaling Proce-Dure for Measuring Mental Workload. In P. A. Hancock & N. Meshkati (Eds.), Human Mental Workload. Amsterdam: Elsevier, 1988; pp. 185–218.

[12] Kurtgün, M. Determination of mental workload with the fuzzy vikor method. Msc thesis, Balikesir University Institute of Science Industrial Engineering. Balıkesir, Turkey.

[13] Riono, R., Suparno, S.& Bandono, A. (2018). Analysis of mental workload with integrating NASA TLX and Fuzzy method. Journal Asro, 9(1), 37-45.

[14] Cooper, M. M., & Kerns, T. S. (2006). Changing the laboratory: Effects of a laboratory course on students' attitudes and perceptions. Journal of Chemical Education, 83(9), 1356.