Lesson Planning with LLMs: AulaNova's Approach to Supporting Instructional Design

Alexandra Sheykina[1], Gabriele De Vito[2], Filomena Ferrucci[3], Fabio Palomba[4], Andrea De Lucia[5]
SeSa Lab - Department of Computer Science, University of Salerno, Fisciano, Italy
[1]asheykina@unisa.it, [2]gadevito@unisa.it, [3]fferrucci@unisa.it, [4]fpalomba@unisa.it, [5]adelucia@unisa.it

Extended Abstract

Context. In recent decades, digital transformation processes have had a significant impact on educational practice. In Italy, as in many other countries, particular attention is paid to the development and implementation of digital technologies in schools and universities. In the daily activity of a modern teacher, characterized by a multiplicity of tasks and responsibilities, the effective use of modern technologies becomes a crucial factor for successful teaching. In the emerging field of ChatGPT for Education, numerous scientific articles have been published that explore the application of pre-trained language models in education. For example, some studies have proposed the use of these models for the automatic generation of quizzes [1], while others have demonstrated their potential in supporting the creation of programming exercises and code explanations [2], as well as in producing content and activities for disciplines such as English, geography, mathematics, science, and medicine [10], [12]. Large language models (LLM) can also assist teachers in lesson planning and in designing inclusive and differentiated learning activities [3]–[8], [13], [15].

Goal: The study aims to reduce the complexity of course design and make it more accessible, faster, easier, technologically advanced, and student-centered in the classroom context. The complexity being addressed refers to the challenges that teachers face when they have to design complete courses and lessons from scratch: identifying competencies and learning outcomes, structuring Units of Learning Activities (UDAs), creating individual lessons, preparing teaching materials such as slides, assessments, and activities, ensuring inclusivity and personalization for students with special needs, and maintaining coherence with educational standards and their own teaching methodologies. The tool does not aim to replace the teacher's role as an educational professional, but to offer structured, customizable suggestions that can be integrated and improved by teachers according to their needs and pedagogical choices.

Motivation. As artificial intelligence increasingly permeates all areas of human activity, it is becoming crucial for teachers to quickly master the new capabilities of this technology. By integrating AI tools into lesson design and classroom activities, educators can leverage these innovations to improve teaching outcomes, adapt to evolving student needs, and foster more engaging and inclusive learning environments.

Novelty. In our work, we treat lesson planning for computer science as an integral part of the educational process within the context of the academic year. We consider several crucial factors, including the classroom context, students' cognitive level, pace of learning, and the presence of students with special educational needs. Moreover, the tool enables teachers to tailor the generated material to the specific needs and teaching methods of their classroom, ensuring flexibility and customization. In particular, teachers can modify the difficulty of the content, integrate various teaching methodologies, and remove or add parts to adapt the materials to specific time constraints and instructional objectives. Notably, to the best of our knowledge, this is the first study focusing on the entire course curriculum and lesson planning for computer science using GPT-4o in secondary school education.

Description of the Aula Nova tool. AulaNova aims to simplify and accelerate the preparation of teaching materials in the computer science field, aligned with educational standards and inclusive practices. Teachers provide essential input, such as subject, school type, and specific classroom context, through free text entries expressed in natural language, with particular attention to the presence of students with special educational needs. The system guides teachers through three main phases: the generation of an instructional design, lesson planning, and the creation of simulated lessons with supporting materials. The tool first generates a detailed teaching plan that includes all the UDA scheduled for the school year. Next, after selecting the relevant UDA, teachers can generate a list of lessons and choose which lesson to plan in detail. The simulated lesson creation phase is divided into two sub-phases. The pre-active phase covers lesson design, including a description of the classroom context, dispensatory measures, and compensatory tools for BES students, involved skills and knowledge, educational objectives, prerequisites, teaching methodologies, any authentic tasks, assessment criteria, and verification methods. The active phase involves using the prepared materials to deliver the lesson, facilitating student interaction, and presenting the topics. Additionally, AulaNova supports the creation of assessment materials, including prerequisite checks and evaluations of lesson objectives in both written and oral form, to reinforce the lesson. Each content can be adapted by the teacher before or after generation with specific feedback. At the end of the process, all materials can be exported in a compressed folder containing the instructional design in Word format, the simulated lesson in PowerPoint format, student

self-assessment diaries, prerequisite checks, and written and oral tests for assessing the achievement of lesson objectives, all in Word format.

Methodology. We recruited N=60 adult teachers enrolled in a 60-credit secondary school computer science teaching qualification course. These participants already hold an academic degree and are developing pedagogical skills needed for certification. As mature, motivated educators, they are familiar with digital tools and open to adopting innovative approaches, such as LLM-based systems, to support their teaching. To evaluate the tool, we designed three post-experiment questionnaires focusing on usability, satisfaction, and technology acceptance. To optimize the performance of the LLM in generating relevant and context-aware outputs, we employed advanced prompt engineering techniques, including persona patterns and output templates, as well as Retrieval-Augmented Generation (RAG) to instruct the model to provide precise and tailored responses. The evaluation criteria and measurement approach were inspired by established practices in the field of software metrics and quality assurance [9], ensuring that the questionnaires were both rigorous and relevant to educational technology, and we followed the *ACM/SIGSOFT Empirical Standards*[1]. We decided to use a Likert Scale with five values [11] for qualitative studies in software engineering. In terms of reporting, we employ the guidelines by Wohlin et al. [14].

Experimental Study in Progress. The preliminary results indicate that the teachers involved in the study had a very positive experience with the tool, reporting high levels of perceived usefulness, productivity, and ease of use. The overall scores were consistently high, with most participants rating the tool between 4 and 5, confirming its value as an intuitive and effective resource for lesson planning. General comments echoed this sentiment, highlighting the well-structured nature of the experiment and the practical advantages of the tool in preparing teaching materials. Teachers particularly appreciated the ease with which they could customize the materials, simply providing natural language feedback, the presence of a structured starting point for planning, and the reduction in cognitive load due to the tool's clean and straightforward interface. Suggestions also emerged to enhance the tool with features for managing user accounts and storing institution-specific teaching materials, further expanding its utility. Overall, the experience highlighted AI's potential as a collaborative teaching partner, providing intuitive, well-organized, and innovative educational resources.

Conclusion and future work. Using GPT-4o in education can significantly reduce lesson planning time and help create inclusive, personalized materials. However, effective use requires carefully composed prompts that consider class context and student learning styles. While offering great potential, GPT-4o also poses risks such as inaccuracies and overly generic content, highlighting the need for careful oversight. Future plans include extending GPT-4o's application to more subjects and educational levels to further support teachers in delivering engaging, personalized learning experiences.

[1]The *ACM/SIGSOFT Empirical Standards*: https://github.com/acmsigsoft/EmpiricalStandards

Acknowledgment

This work was partially supported by project FAIR (PE0000013) under the NRRP MUR program funded by the EU - NGEU.

References

[1] R. Dijkstra, Z. Genç, S. Kayal, J. Kamps, et al., "Reading Comprehension Quiz Generation using Generative Pre-trained Transformers," in *iTextbooks@ AIED*, 2022, pp. 4–17.

[2] S. Sarsa, P. Denny, A. Hellas, and J. Leinonen, "Automatic generation of programming exercises and code explanations using large language models," in *Proceedings of the 2022 ACM Conference on International Computing Education Research – Volume 1*, 2022, pp. 27–43.

[3] B. Hu, L. Zheng, J. Zhu, L. Ding, Y. Wang, and X. Gu, "Teaching plan generation and evaluation with GPT-4: Unleashing the potential of LLM in instructional design," *IEEE Transactions on Learning Technologies*, 2024.

[4] B. Hu, J. Zhu, Y. Pei, and X. Gu, "Exploring the potential of LLM to enhance teaching plans through teaching simulation," *npj Science of Learning*, vol. 10, no. 1, p. 7, 2025.

[5] A. Goslen, Y. J. Kim, J. Rowe, and J. Lester, "LLM-based student plan generation for adaptive scaffolding in game-based learning environments," *International Journal of Artificial Intelligence in Education*, pp. 1–26, 2024.

[6] H. Fan, G. Chen, X. Wang, and Z. Peng, "LessonPlanner: Assisting novice teachers to prepare pedagogy-driven lesson plans with large language models," in *Proc. 37th Annu. ACM Symp. User Interface Softw. Technol.*, pp. 1–20, 2024.

[7] Y. Zheng, X. Li, Y. Huang, Q. Liang, T. Guo, M. Hou, B. Gao, M. Tian, Z. Liu, and W. Luo, "Automatic lesson plan generation via large language models with self-critique prompting," in *Proc. Int. Conf. Artif. Intell. Educ.*, pp. 163–178, 2024.

[8] M. Moundridou, N. Matzakos, and S. Doukakis, "Generative AI tools as educators' assistants: Designing and implementing inquiry-based lesson plans," *Computers and Education: Artificial Intelligence*, vol. 7, p. 100277, 2024.

[9] N. Fenton and J. Bieman, *Software Metrics: A Rigorous and Practical Approach*. CRC Press, 2014.

[10] C. K. Lo, "What is the impact of ChatGPT on education? A rapid review of the literature," *Education Sciences*, vol. 13, no. 4, p. 410, 2023.

[11] B. A. Kitchenham and S. L. Pfleeger, "Personal opinion surveys," in *Guide to Advanced Empirical Software Engineering*, Springer, 2008, pp. 63–92.

[12] N. Rutten, W. R. Van Joolingen e J. T. Van Der Veen, "The learning effects of computer simulations in science education," *Computers & Education*, vol. 58, n. 1, pp. 136–153, 2012.

[13] R. O. Davis and Y. J. Lee, "Prompt: ChatGPT, create my course, please!," *Education Sciences*, vol. 14, no. 1, p. 24, 2023.

[14] C. Wohlin, M. Höst, and K. Henningsson, "Empirical research methods in software engineering," in *Empirical Methods and Studies in Software Engineering: Experiences from ESERNET*, Springer, 2003, pp. 7–23.

[15] Geesje van den Berg and Elize du Plessis, "ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education," *Education Sciences*, vol. 13, no. 10, pp. 998, 2023.