

Chapter # 21

“DOES IT SAVE TIME AND MAKE YOU SMARTER?” – LESSONS LEARNED FROM A WORKSHOP DEVELOPING TEACHERS’ PROMPTING COMPETENCE WITH CHATGPT

Björn Sjödén¹ & Jeroen Lievens²

¹School of Education, Halmstad University, Sweden

²Faculty of Engineering Technology, University of Leuven, Belgium

ABSTRACT

The use of generative AI (GenAI) such as ChatGPT highlights the need for educators to develop novel critical competencies. We present a process-oriented approach for developing teachers’ prompting competence – the ability to formulate effective instructions to GenAI tools – and our first evaluation of its implementation in a workshop for teachers in higher education. We argue for an “anti-shortcut” approach, meaning that effective AI use requires slowing down, reflecting, and collaboratively refining one’s interactions to reach desirable outcomes. 15 teachers participated in the workshop, which focused on a typical professional task (assessing student essays) using stepwise prompts in ChatGPT. The participants were instructed to use ChatGPT or similar GenAI tool to grade a short student essay, using a scoring rubric, and to write some feedback to the student. The GenAI output was then compared to a human professional assessment of the same text. As part of developing the teachers’ skills, we introduced the strategies “prompt chaining” and “few-shot prompting”. The participants were asked to revise and improve their prompts, so that the GenAI assessment matched that of the human as closely as possible. The results fueled a discussion on teachers’ use of GenAI, from which we report some key insights.

Keywords: generative AI, GenAI, teacher training, higher education, ChatGPT, prompting.

1. INTRODUCTION

There is no doubt that generative artificial intelligence, or GenAI, is drastically changing educational practices. However, users are still in the adaptation phase of embracing the technology – some with more hesitation than others, partly reflected in the title quote by one teacher – and it is becoming equally apparent that the extent and types of uses of GenAI vary widely between different academic institutions, as well as between individual teachers and students within the same institution. The academic literature can hardly keep up to date with this development. Some practice-oriented books quickly followed the unprecedented success of ChatGPT 3.5 (e.g. Hunter, 2023; as the author notes in the introduction, ChatGPT took five days to reach one million users; Instagram took 2.5 months and Netflix took 14 months).

An emerging core competence is so-called “prompt engineering” – that is, how instructions to the GenAI can be effectively formulated and improved – which inspired our present proposal for increasing teachers’ prompting competence in relation to student work (which itself may involve using GenAI). Considering that presented examples of outcomes and limitations may become outdated already at the time of publication as the technology advances, we hold that any prescribed guidelines should be combined with actual training

sessions to maintain their relevance and applicability. Therefore, rather than presenting empirical findings from our own work, or a ready-made recipe for teachers to follow, we chose a process-oriented approach in the form of a workshop, which aims to develop the participants' current knowledge of the present technology (in this case, the most used version of ChatGPT) in relation to a benchmark set by human professionals.

In this chapter, we present the outline of such a workshop, and its first implementation with teachers in higher education. There were two points we found particularly worth considering in preparation. First, we hold that human teachers remain the experts in their field, but that they may learn how to use GenAI, such as ChatGPT, for carrying out certain educational tasks, such as assessing a student text, more effectively, while maintaining their professional standards and integrity. To achieve this, teachers must learn how to prompt properly and gauge GenAI output judiciously. Second, we must consider the pitfalls associated with achieving deceptively quick results with technology that could mislead users to taking shortcuts and by-pass their critical thinking (Zhai et al., 2024). In other words, we must think before we act. The idea of a “collective anti-shortcut approach” warrants some elaboration before we describe its actual implementation and lessons learned from the workshop.

2. BACKGROUND

2.1. Rationale: The Problem of Shortcuts

Humans are wired to seek efficiency. In cognitive psychology, this tendency manifests as the use of heuristics, in the form of mental shortcuts that save time and effort. We believe that people – teachers and students included – are entitled to take shortcuts in known territories, in order to reach their goals faster and in the service of others. For students, this could mean using ChatGPT for explaining a concept they did not understand from the teacher, rather than having to ask the teacher again (if at all possible). For teachers, it could mean using ChatGPT for constructing or making an assessment, to free up their time and resources for helping students learn rather than just assessing them. In both cases, it is assumed that the user already has the required skill (asking for help, constructing a test) which has been preceded by much longer learning (going to school, taking a research degree). Importantly, what is being produced (knowledge) must not be confused with the process of production (learning).

Whereas productivity gains with GenAI can be quick, human learning is slow and requires time for reflection and adaptation. Taking shortcuts using AI can therefore become epistemic traps, bypassing critical thinking, verification, and creativity. There is overwhelming evidence that humans resist mental effort when they can (Willingham, 2009), and that hard thinking is generally associated with negative affect (David et al., 2024). This fact itself may explain what Zhai et al. (2024) refer to as users' tendency to take “cognitive shortcuts”, or heuristics, when using AI, in the face of ethical issues and at the risk of a negative impact on their cognitive abilities.

A core risk that applies to students and teachers alike is an over-reliance on the AI output, without proper questioning and verification (including factors such as AI hallucination, algorithmic bias, plagiarism, privacy concerns, and transparency issues, and how over-reliance impacts decision-making, critical thinking, and analytical thinking). In short, taking shortcuts with AI may lower educational quality (Ahmad et al., 2023), unless careful attention is given to such issues. In our case, we specifically target how teachers may avoid taking invalidated shortcuts to assessment by carefully examining and reconstructing their prompts, while acknowledging that AI can be highly useful for this kind of task.

For example, good prompt engineering has been highlighted as critical to avoid and mitigate the weaknesses of using GenAI, with reference to ChatGPT specifically (Johnson, 2024).

Hence, in order to mitigate the negative effects of taking delusive shortcuts, we presented the idea of taking an “anti-shortcut” approach for teachers to learn better prompting (Sjödén & Lievens, 2025). When training prompting skills in a group, we considered that users would benefit from deliberately slowing down and reflect upon both the specific formulation of prompts and their generated output. This would entail first engaging together in possibly effortful but rewarding thinking, before actual and productive shortcuts can be created to facilitate teachers’ work.

2.2. Related Empirical Work and Present Educational Needs

A growing body of empirical research investigates the pedagogical implications of GenAI, like ChatGPT, and the large language models (LLMs) they are based on. While several studies emphasize its potential to foster critical thinking and analytical skills when integrated thoughtfully into learning designs (Guo & Lee, 2023; Dobre & Popescu, 2024), others highlight significant risks, particularly when these tools are used without proper scaffolding.

A recent preprint study by researchers at the MIT Media Lab raises important concerns about the cognitive cost of relying on ChatGPT. In the study, Kosmyna et al. (2025) asked participants to complete SAT-style essay tasks under three conditions: unaided (“brain-only”), using a search engine, or using ChatGPT. The researchers tracked participants’ brain activity using EEG and found that those in the ChatGPT condition showed significantly lower levels of neural connectivity (an indication of reduced mental engagement) compared to those using their own cognitive resources or traditional search. These participants also had weaker recall of their writing and reported lower feelings of authorship. Over time, their essays became increasingly homogenous and reliant on copy-paste behavior, leading the researchers to suggest that unreflective use of LLMs may incur a “cognitive debt” that undermines deeper learning and original thought. Although the study has not yet been peer-reviewed, it provides early empirical support for the hypothesis that LLMs, if not critically integrated into pedagogy, may erode the very cognitive skills they are meant to support.

Beyond cognitive engagement, another key concern is the widespread misunderstanding of how LLMs function. Students and educators often interpret LLM responses as factually grounded or authoritative, even when they are merely plausible-seeming predictions based on training data (Woodland, 2023; Yang, 2024). This unpredictability can be attributed to the probabilistic mechanisms underpinning generative models, yet many users assume deterministic behavior akin to earlier rule-based software systems.

Given these findings, it becomes clear that educational institutions must not only train students in how to use LLMs but also develop their understanding of how these models work. This includes fostering critical awareness of output variability, potential bias, and the limits of LLM-generated knowledge. Aligning ourselves with, e.g. Royer (2024), we argue that such AI literacy should be extended to faculty and staff as well, particularly in the light of growing institutional reliance on generative AI for teaching, assessment, and communication.

3. WORKSHOP DESIGN AND IMPLEMENTATION

Our overarching goal was to provide a structured training opportunity for teachers that targeted how they could avoid taking invalidated shortcuts to assessment, by carefully examining and reconstructing their own prompts. This seemed particularly apt for a workshop format, which included practice and perspective-taking, unlike what could be achieved by users following instructions from a book or article on their own. Next, we elaborate on the requirements for carrying out effective workshop tasks and how these were carried out in practice.

3.1. Developing Teachers' Prompting Competence

Prompting competence refers to the ability to design, test, and refine instructions to an AI model in ways that yield valid, transparent, and educationally meaningful outputs. In this context, the workshop participants were teachers who were guided through a task simulating authentic assessment of a student essay. Each group was asked to prompt ChatGPT to grade a student text and provide feedback, using a scoring rubric familiar to most higher education teachers. Our aim was not only to compare GenAI results with human grading but also to examine how prompt phrasing influenced the AI's reasoning.

Referring to what we further labeled a “collective anti-shortcut design”, we assigned participants to work in small teams, iteratively revising their prompts, observing the AI's responses, and reflecting on what made certain prompts more effective. This exercise served to reveal that prompting is a form of pedagogical reasoning, which requires hypothesis testing, analytical thinking, and ethical sensitivity.

3.2. Techniques for Improving Prompts

Recent research highlights two particularly effective methods for improving AI prompting: few-shot prompting and prompt chaining (Gilardi et al., 2023). These methods help educators structure their interactions with AI tools more deliberately, reducing superficiality and bias.

Few-shot prompting involves showing the AI a few examples of good and poor outputs before it performs a new task. This contextual “training” helps align the AI's reasoning with human expectations. For instance, before grading a student essay, a teacher might provide examples of what counts as “good use of theory” and “weak use of theory”, so that ChatGPT learns from these standards.

Prompt chaining involves decomposing a complex task into smaller, sequential steps. For example, when assessing a student essay, the first prompt might ask ChatGPT to identify the essay's sections, the second to summarize each section's argument, and the third to evaluate the text against given criteria such as theoretical use or coherence. Each stage builds on the previous one, helping the AI avoid generic responses and enabling the teacher to monitor reasoning transparency.

Combining few-shot prompting and prompt chaining – starting with few-shot examples and then chaining prompts step-by-step – may significantly improve reliability. This hybrid method trains the AI to understand evaluative criteria while guiding it through a transparent reasoning process.

3.3. Workshop Structure and Procedure

The workshop was designed as a 45-minute, hands-on session combining brief instruction, collaborative experimentation, and collective reflection. A preliminary outline of the workshop was presented in a previous publication (Sjödén & Lievens, 2025). Here follows a refined description in retrospect, with respect to how the workshop was carried out in practice with the participating teachers.

The workshop took place as part of the END conference in Budapest, June 2025. The participants were 15 teachers in higher education (all conference attendees) who joined the session as part of the conference program. The session began with a short introduction outlining the concept of prompting competence and why it matters for higher education. The two facilitators (the present authors) emphasized that while generative AI can accelerate many routine academic tasks, its pedagogical use requires critical awareness to avoid over-reliance and unreflective “cognitive shortcuts”. Participants were then introduced to the challenge: to use ChatGPT (or similar tool) to assess a short student essay as accurately and transparently as a professional human grader (an anonymous university teacher). They were provided with a (slightly revised and anonymized) student essay and a scoring rubric; their task was formulated as “Prompt the AI to: 1) suggest a grade, 2) write a comment of 5-6 lines with feedback to the student”.

Having worked in small teams (typically groups of three) with the task, the various outputs from ChatGPT were compared to an actual, human assessment of the same text (which had been done beforehand), and the differences were discussed.

Next, participants were introduced to the two prompting improvement strategies described above, prompt chaining and few-shot prompting. The same groups were asked to iteratively refine their prompts and experiment with wording, sequencing, and follow-up instructions to improve the AI’s output. Their renewed task was formulated as, “Can you specify your prompts such that the output assessment by the AI as closely as possible reaches that of the human benchmark?”.

Finally, the workshop concluded with a reflection on lessons learned, and what made some prompts work better than others. Participants discussed how deliberate prompting could improve not only AI output quality but also their own analytical and pedagogical thinking. The facilitators summarized the key takeaways.

4. RESULTS AND EVALUATION

A first notable finding – to the surprise of the workshop participants – was that the same prompt for the exact same essay using the exact same GenAI tool yielded divergent responses, ranging from weak (fail) to excellent (high pass). In our interpretation, because almost all software before LLMs was basically hard-coded, it was hard for the participants, and for teachers in general, to conceive computer-generated output as probabilistic rather than deterministic (and therefore, to a degree, random). We hold that this deeper insight into the nature of LLMs instill users with the critical attitude that is necessary for, ultimately, using such tools productively and judiciously. Therefore, we maintain that a) faculty needs to be trained for using these novel technologies in the domain of education and b) this training, rather than limiting itself to a “tips and tricks” approach, should also explore the fundamental *modus operandi* of LLM technology.

A second surprising result for the workshop participants was that, on average, the GenAI assessment tended to be milder – by a significant degree, in some cases – than the human assessment. In fact, the human assessor failed the essay that the workshop participants had submitted to the GenAI tool. By contrast, one of the GenAI tools that the workshop

participants used, awarded the same essay with the highest possible grade (“Pass with distinction”). It was only after being reminded of one the rubric’s evaluation criteria – that two theories needed to be applied in the essay – that the GenAI tool readjusted its initial assessment. This experience helped workshop participants understand the importance of good prompting techniques. Figure 1 shows the respective assessments of the GenAI tool (in the present case ChatGPT) and the human assessor, including their feedback to the student.

Figure 1.
Assessment of a student essay by ChatGPT as compared to an assessment of the same text by a human university teacher.

Initial assessment by ChatGPT: Pass with distinction (VG).

A well-structured and interesting essay with a good foundation in both theory and practice. Theoretical perspectives and scientific research are clearly integrated. Language and form are generally acceptable but need to be worked on linguistically in a few areas. The text could benefit from further deepening of the critical and problematizing reasoning.

Human assessment: FAIL

The essay lacks systematics, analysis and adequate use of concepts. The essay does not adequately apply two theories and/or perspectives from the course literature. It does not follow the instructions of the assignment; go back to the guidelines and check that all parts of the assignment are addressed. The text has good readability, but take note of comments and markings in the text to increase the academic quality of the work.

A third result to emerge from the workshop was that the implementation of proper prompting techniques, such as the above-mentioned “chain prompting” and “one-shot-prompting”, did significantly improve the GenAI tool’s evaluation in terms of aligning with the human assessor. When the prompt was broken down into simpler and consecutive steps, the GenAI tool was more fastidious in applying the rubric criteria. Likewise, the assessment quality improved when the tool was fed with examples of essay and human-made assessments of these essays. The assessments that matched the human-made assessments most closely were those that emerged from a hybrid prompting approach, combining chain prompting and one-shot prompting.

Finally, the workshop participants remarked that, even when using sound prompting techniques, the LLM’s still needed human oversight. Critical follow-up questions were necessary to ensure that the pedagogical standards that the participants set for themselves were met. It was broadly concluded that while the GenAI tool can offer valuable support, it cannot be relied upon to conduct assessment independently, even when given carefully crafted prompts.

5. DISCUSSION

A key insight from our workshop was the widespread misunderstanding among the participants of how large language models in tools like ChatGPT actually function. Many educators and participants appeared unfamiliar with the idea that LLM outputs are generated probabilistically rather than deterministically. This lack of foundational knowledge often led to confusion when identical prompts produced highly divergent responses, from weak to exemplary. Future research should therefore explore how teachers’ and other users’ conceptual understanding of LLM architecture – particularly its non-deterministic, stochastic nature – influences their prompting behavior, their trust in AI-generated responses, and ability to critically assess the quality of output.

In the context of education, we maintain that shortcuts are not inherently negative. Teachers routinely use templates, rubrics, and automated systems to increase efficiency for planning lessons and assessing student work. Issues arise when shortcuts become substitutes for thinking. The “anti-shortcut” approach proposed here is not about rejecting efficiency or use of GenAI tools, but about ensuring that it serves pedagogical integrity and quality by making time for reflection.

The workshop results suggest that developing prompting competence is not primarily a technical challenge but a pedagogical and ethical one. We deem that teacher education programs should integrate GenAI training not as an add-on, but as a form of professional reasoning practice, where collaborative reflection on multiple examples (which allow for comparisons) makes a critical factor. By engaging teachers in “reverse engineering” of successful prompts, the logic of good AI interaction can be made visible. This method can foster teachers’ further professional development, as they continually learn to adapt their prompting strategies to new models and contexts.

6. FUTURE RESEARCH DIRECTIONS

Future studies should examine the role of “AI epistemic literacy” more broadly, in both student and faculty populations. While existing research has begun to document misconceptions and overreliance on LLMs (Yang, 2024; Woodland, 2023), little is known about how targeted interventions that explain the internal logic of LLMs affect educational outcomes. Experimental studies could assess whether brief instruction on LLM mechanisms enhances skills related to prompting competence, critical thinking, and metacognitive monitoring. Additionally, longitudinal research could track whether repeated, uncritical use of GenAI leads to cognitive offloading or erosion of independent reasoning skills. As institutions formulate GenAI policies, it seems pertinent that future work also evaluate which strategies most effectively support responsible and informed use, with respect to staff training and curriculum integration.

7. CONCLUSION

Developing prompting competence is essential for ensuring that AI supports, rather than supplants, human learning. Our collective anti-shortcut approach emphasized deliberate, reflective use of GenAI: slowing down to think, testing prompts collaboratively, and aligning AI use with educational values. Techniques like prompt chaining and few-shot prompting exemplify how structured interaction can enhance both AI reliability and teacher understanding. Ultimately, prompting competence is not just about mastering a tool; it is

about cultivating a mindset of critical curiosity, where the GenAI tool is treated as a partner in inquiry rather than a shortcut to definite answers.

In sum, our message to teachers from trying this out with the target group in practice, is threefold. First and foremost, never trust the GenAI's first response and be cautious of bias (which is often positive and people-pleasing). Second, refine your prompts, such as by chaining and few-shot prompting. Third, ask for explicit justifications and follow-up questions to ensure coherence and validity. Our experience is that carrying these tasks out in the form of a workshop provided concrete, added value to the participants – not the least judging by their surprise at some of the results – beyond what would have been achieved by simply reading about GenAI and LLMs in a paper or online manual.

Our overall conclusion might serve as a response to the question by the participant quoted in the title of this chapter: Using GenAI might *not* save you time, but with proper prompting competence and your attention to its peculiarities discussed in this chapter, it could well increase the quality and consistency of your work, in the *same* amount of time.

REFERENCES

- Ahmad, S. F., Han, H., Alam, M. M., Rehmat, M., Irshad, M., Arra.o-Mu.oz, M., & Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications*, 10(1), 1-14.
- David, L., Vassena, E., & Bijleveld, E. (2024). The unpleasantness of thinking: A meta-analytic review of the association between mental effort and negative affect. *Psychological Bulletin*, 150(9), 1070-1093.
- Dobre, S.-C., & Popescu, E. (2024). Exploring students' perception and experience with ChatGPT and critical thinking in a higher education context. *2024 21st International Conference on Information Technology Based Higher Education and Training (ITHET)*, 1-6.
- Guo, Y., & Lee, D. (2023). Leveraging ChatGPT for enhancing critical thinking skills. *Journal of Chemical Education*, 100(12), 4876-4883 Retrieved from https://pubs.acs.org/doi/pdf/10.1021/acs.jchemed.3c00505?ref=article_openPDF
- Hunter, N. (2023). *The Art of Prompt Engineering with ChatGPT*. Nathan Hunter: Independently published.
- Johnson, W. L. (2024). How to harness generative AI to accelerate human learning. *International Journal of Artificial Intelligence in Education*, 34(3), 1287-1291.
- Kosmyna, N., Bartley, J. E., Khan, R., & Picard, R. W. (2025). *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task* [Preprint]. MIT Media Lab. Retrieved October 23, 2025, from <https://www.media.mit.edu/projects/your-brain-on-chatgpt/overview/>
- Royer, C. (2024). Outsourcing humanity? ChatGPT, critical thinking, and the crisis in higher education. *Studies in Philosophy and Education*, 43, 479-497.
- Sjöden, B., & Lievens, J. (2025). Developing Teachers' prompting Competence Using ChatGPT: A Collective Anti-Shortcut Approach. *Education and New Developments (END) 2025* (Vol. 2, pp. 662-664). InScience Press.
- Woodland, L. (2023). ChatGPT for improving medical education: Proceed with caution. *Mayo Clinic Proceedings: Digital Health*, 1(3), 294-295.
- Yang, Y. (2024). Towards responsible use: Student perspectives on ChatGPT in university education. *Proceedings of the 23rd European Conference on e-Learning - ECEL 2024*, 23(1), 415-422. Retrieved from <https://papers.academic-conferences.org/index.php/ecel/article/view/2790>
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments*, 11(1), 28.

ACKNOWLEDGEMENTS

The GenAI tool ChatGPT by Open AI served as a valuable assistant, supporting the initial drafting and writing process of this paper. All GenAI output has been thoroughly reviewed and revised before selected content was integrated into the full text. The authors assume complete responsibility for the final manuscript.

AUTHORS’ INFORMATION

Full name: Björn Sjöden

Institutional affiliation: School of Education, Halmstad University, Sweden

Institutional address: Box 823, 301 18 Halmstad, Sweden

Biographical sketch: Björn Sjöden is Assistant Professor in Educational Science at Halmstad University with a Ph.D. in Cognitive Science. His research focuses on how digital and analogue tools can be combined to optimize teaching and learning processes, informed by the science of learning.

Full name: Jeroen Lievens

Institutional affiliation: Faculty of Engineering Technology, University of Leuven, Belgium

Institutional address: Campuslaan, 3590 Diepenbeek, Belgium

Biographical sketch: Jeroen Lievens is Associate Professor at ETHER (Engineering Technology Education Research), KU Leuven, Campus Diepenbeek. He teaches communication and research skills to engineering students. His research focuses on student-driven action research, investigating the workplace relevance of communication and GenAI competencies for engineers.