This document is part of the Erasmus+
project 2024-1-NO01-KA210-ADU000244898

THE GLOW GUIDEBOOK

Generative AI Prompt Writing for Language Teachers

Co-funded by
the European Union

Introduction

This guidebook aims to empower adult language teachers with practical knowledge and skills in using generative AI tools for lesson planning. It is a direct output of the Erasmus+ project "Generative AI: How to Write Prompts for Language Teachers: Easing the workload in the classroom" (Project Acronym: GAHWP-LT), realized by Briga AS (Norway) and Katholiek Onderwijs Vlaanderen (Belgium). The guidebook is designed to reduce teacher workload, address the diversity of learners, and foster confidence and competence in digital transformation.

Welcome to the GLOW guidebook, an Erasmus+ KA210 funded project which has aimed to increase the knowledge and use of generative AI in second language classrooms. Two organisations, Briga (Norway) and Katholiek Onderwijs Vlaanderen (Belgium) have come together, with their teachers, to test the use of Generative AI in creating lesson plans, with an overarching aim of decreasing the workload of our teachers in an increasingly demanding sector. The guidebook/webpage aims to serve as a helper in how to best use generative AI in the lesson planning phase, and how to get the best results from your prompts, whilst also saving time and decreasing the cognitive load on teachers. Originally called *"Generative AI: How to Write Prompts for Language Teachers: Easing the workload in the classroom" (Project Acronym: GAHWP-LT).*

The guidebook is split into 6 chapters with the aim of taking you through the entirety of the project. Briefly summarized, chapters 1 to 3 are related to the background of the project and the background related to generative AI. In these chapters we cover the motivation behind the project, what and where Generative AI is and how it has developed and lastly, we focus on our teachers and how they gave us feedback which helped us define the project further. Chapters 4 to 6 are practical chapters, which will provide a more hands-on experience on generative AI. These chapters are the result of over 200 lesson plan submissions by teachers who are teaching Norwegian or Dutch as a second language, with analysis and evaluations conducted by the teachers and the project team throughout the project. These sections aim to assist you in using generative AI for lesson planning and will be the fundamental pillars of adopting the results of this project into your work. We conclude the guide with chapter 7 and appendices, dedicated to further reading and practical documents, as well as conclusive statements.

Chapter 1. About This Guidebook

This section sets the context for the guidebook, describing the motivation for its creation, its intended users, and a roadmap for navigating its contents.

In the following chapter we will cover the project background and its objectives. The partner organisations will be presented along with their roles. We will clarify who our target audience is, and what their role is and finally we will summarise how to use this guidebook.

Project Background

The GLOW project was inspired in part by teacher complaints and partly by a desire for professional development by pedagogical leaders. In the wake of the Russian invasion of Ukraine, European second language classrooms were faced with a new group of learned in a scale that they were unprepared for. Similar events in the wake of the Arab spring and the inclusion of Eastern-Europe into the EU have spurred teachers to adapt. In this most recent influx, the technological advancements within machine learning had led to generative artificial intelligence. The adoption of generative AI was quickly deemed pivotal for survival, regardless of which field you belonged to. Simultaneously, teachers faced an incredibly challenging situation in having to adapt to varying learners, all coming into their classroom, with different needs and requiring more planning time than normal. The partners therefore decided to explore the idea of using generative AI to solve the issue of workload for teachers.

Motivated by several other aspects, such as increasing digital skills as well as allowing teachers to retain autonomy, the GLOW project wanted to give teachers a tool, but also a guide in how to use the tool. To avoid everything being a nail, and generative AI being the metaphorical hammer, the project strived to teach and prove that generative AI could lessen the workload of teachers in their challenging classrooms.

Teachers in both organisations had varying degrees of experience with the use of generative AI professionally, and a spectrum of feelings regarding generative AI. With extreme ideas on both ends of the spectrum, such as “gen AI is useless” to “gen AI will take my job”, the project group decided to ensure a basic foundational understanding for all teachers. Three activities were set up, to ensure that teachers were involved continuously and that they were represented. Firstly, teachers were given a chance to answer a questionnaire regarding their challenges and knowledge regarding generative AI. Secondly, they received training based on this and were given the task of producing 10 lesson plans over a year, with two more trainings provided to them during this period. Finally, they will see the guide and influence its content to ensure that is helpful

and easy to adopt. Training the teachers was essential for the project to create a minimum knowledge foundation for all participants.

GLOW has the underlying aim to create digital resilience amongst its test group and to expand to the resilience of all teachers in a rapidly changing environment. To exemplify this, a factor which was outside of the groups control truly visualised this rapid development. At the time of writing the application the choice was made to use ChatGPT, as the most commonly known LLM at the time of writing the application. At the time, version 3.5 was available. In the time since the application was submitted, the project started and finally finished, several new versions have come to market and at the time of writing we are using version 5.2 of ChatGPT. In this we find the wider reason why this project is necessary for teachers. We are in a time where development is rapid, meaning that the skill of adoption and adaptation matters more than singular tools. Regardless of whether our metaphorical hammer is made of titanium, has a carbon fibre handle or comes with a spring loaded force detector with Bluetooth, it is the act of hitting the nail on the head which is important.

The Partner organisations

GLOW is a KA210 project between two organisations, Briga AS from Norway and Katholiek Onderwijs Vlaanderen from Belgium (Dutch speaking). The organisations have established contact at a TCA seminar where they did not originally apply for a project together. In 2023 contact was established again to propose a new project, with Briga AS as the coordinator. Together with pedagogical leaders from KOV, an agreement was quickly reached in how to approach such a project and together the application was successful. Below is a brief description of the two partners.

Briga AS

Briga AS was founded in 2015 with the aim of applying to public tenders which were put to private providers by the Labour and Welfare organisation of Norway. In 2016 they won their first tender, providing assistance to job seekers. In 2016 they also began with Norwegian training on site for Norwegian businesses. This became their main activity from 2020 and onwards and has expanded to many businesses. Briga has been a pioneer during COVID-19 in ensuring that their learners could quickly adapt to an online classroom. This has led to participation in Erasmus+ projects related to developments within language training and hybrid classrooms. Since 2025 Briga has been working on a public tender to help Ukranian refugees learn Norwegian within the hotel, restaurant and catering sector and to find them work as a part of a wider political shift in refugee policy. Briga has many teachers who teach Norwegian as a second language working as part time employees or as self-employed. These teachers have contributed to this project.

Katholiek Onderwijs Vlaanderen (aka Catholic Education Flanders)

Catholic Education Flanders (Katholiek Onderwijs Vlaanderen) is the largest educational network in Flanders and the Brussels region. It is a membership organisation that unites Catholic educational institutions across all levels, including pre-primary, primary, secondary, vocational education (VET), adult education, higher education, and the University of Leuven.

The organisation represents and supports more than 2,200 schools and 935,000 learners, making it a key player in the Flemish educational landscape.

Its mission is to advocate for school boards, support educational innovation, and strengthen quality development through pedagogical, administrative, and managerial guidance.

Project's target audience

From the beginning, the main audience has been teachers, and this project has focused on second language teachers in Norway and Belgium. Whilst we understand that this can feel like a small and exclusive group, we believe that the majority of this guidebook and project can be useful for all professionals who wish to make better use of generative AI in their work.

GLOWs aims are directly linked to a challenge the teachers have been facing, which has been the increased workload in lesson planning. We have therefore included examples and best practice related to their field. The underlying principles however can be transferred to other fields, and we hope that firstly all teachers, of all age groups can make use of this guidebook. Our secondary target groups are learners of second languages themselves. This guidebook will also allow them to create personalised lesson plans, based on their challenges. This was not originally an intended result, but as our teachers make use of generative AI lesson plans, we see that the ease of access allows all users to work in a similar way. Teachers have used their expertise to tweak and adjust some of the results given to them by the generative AI, and we will therefore advise caution when using it for personal use.

For second language teachers of adults, we hope that this guidebook solves the very real problem of time and cognitive constraints. We have observed and received feedback regarding the heavy load teachers face in the newer classrooms and have aimed to lessen this load. The guide is not a complete list of possible ways to use generative AI and nor is it going to be something that can ever be seen as finished. As mentioned in the introduction, in such a rapidly developing field there is room for changes, developments, new ideas and new ways of working. Our guide serves as a foundation to build further upon, a step towards the metaphorical top.

How to use this guidebook and why

The guidebook is split into chapters which clearly state what they can do for you, and how they can aid you in both lesson planning and your professional development. Below is a brief explanation of each chapter.

The Chapters and their uses

Chapter 2: In this chapter you will begin to understand the foundational aspects of generative AI, of LLMs and how they are built. Furthermore, you will be given a brief introduction to the EU AI Act. This chapter aims to educate you on what generative AI is, and serves as a knowledge foundation. Here we hope to boost your confidence in yourself, but also your understanding of the tool you are using.

Chapter 3: If you wish to understand your peers who participated in this project we recommend reading this chapter, as it outlines the background our teachers went in with when starting the project. We present their challenges as well as highlighting what solutions have come forward as a part of the project.

Chapter 4: This chapter is a presentation of the results from the project. You can find the data and the hypotheses from both organisations, as well as a comparison of the data.

Chapter 5: Mandatory reading for using the project results. We go into usage areas of generative AI, with a focus on best practices which go beyond the prompt writing aspect of the project. These are developments and trainings the teachers have received during the project and we wish to share with you.

Chapter 6: If you feel sceptical about AI, and its role in your work we recommend this chapter. You will see the results of the teachers after a year of being involved in the project. We show how teachers have changed their working behaviours, as well as presenting concrete data driven results about how teachers developed and became users of generative AI to decrease their workload.

Why you should use this guide

There are those who would assume a project such as GLOW could lead to the decline of the teacher's role and importance in the classroom. To this, we say "No!".

The GLOW project aims to empower the teachers who are standing in the frontlines of education. Teachers who are working in classrooms where the challenges are a hinder to their excellence. We want to ensure that these challenges and the strain from planning around these challenges is reduced. GLOW aims to keep the teacher in the

classroom, but to create an easier path from day 1, ensuring that improvement does not come at the cost of your mental health or private time.

Furthermore, this guide opens up new possibilities for creativity and to test new ideas in your classroom, ensuring that the first few steps can be taken together with generative AI. We see this guide as releasing many teachers from time constraints related to classic teaching, and hope that the time created from using generative AI can give room for creativity, rest and also new approaches.

As well as idealistic aims, you are securing your position by constantly developing upon tried and tested methods and content. The limitations are only existent if you as a user create them, we see endless possibilities.

Chapter 2. Understanding Generative AI and prompting

The EU AI Act

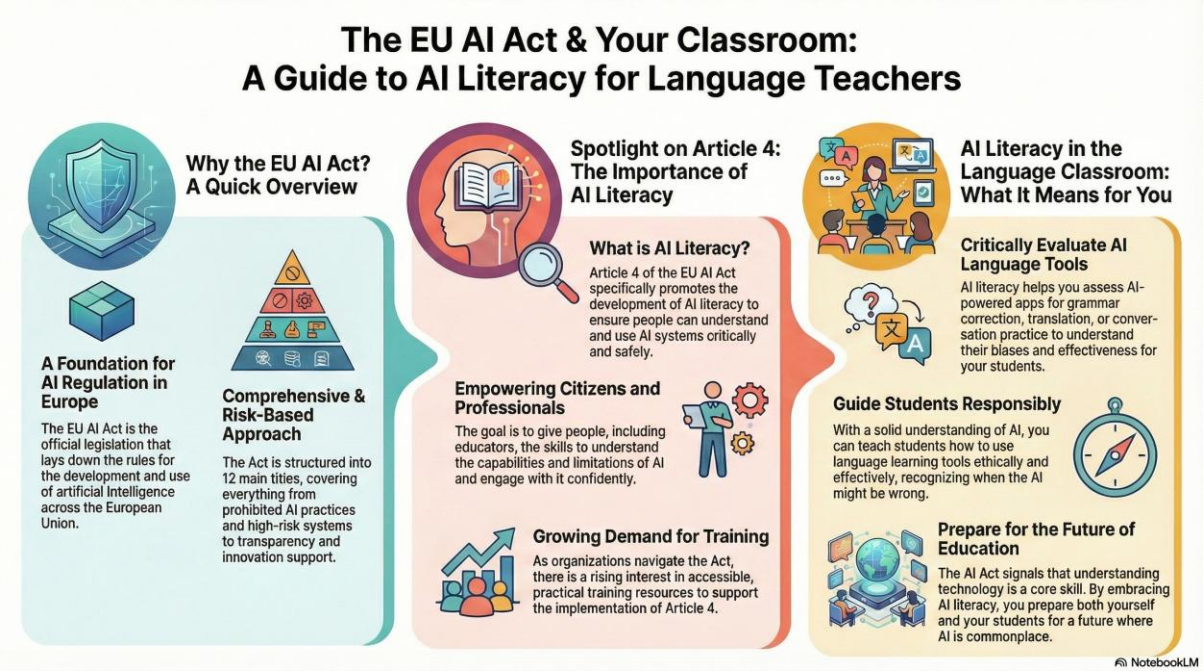
In 2024, the European Union developed the EU AI Act. The core objective of the EU AI Act is simple but crucial: to create a clear framework and unambiguous rules for the development and use of artificial intelligence. It lays the foundation for AI regulation in the EU.

A key article of this EU AI Act within this Erasmus+ KA2 project is Article 4 on AI literacy. AI literacy essentially means promoting a basic level of knowledge and skills about AI among all citizens. It's about people understanding what AI is, how it works, what its possibilities are, and what its risks are. Article 4 can be compared to obtaining a theory certificate: before you drive a vehicle (an AI system), the law requires you to first understand the rules of the road and the basic operation of the vehicle, so you don't pose a danger to yourself or others. Article 4 describes:

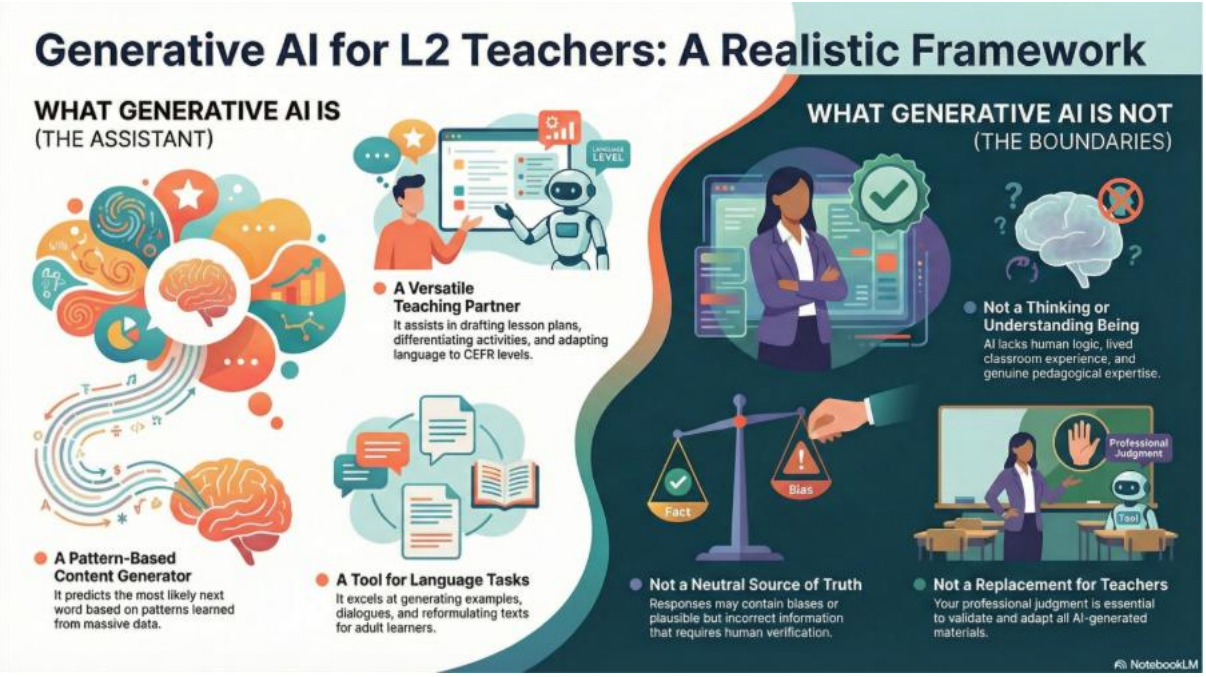
- The importance of knowledge:
  This article emphasizes the importance of organizations and individuals working with AI having sufficient knowledge to understand the systems and associated regulations.
- Implementation support:
  It serves as a basis for organizations to find practical training and resources to properly implement the AI Act.
- Making training accessible:
  This article has led to a growing focus on providing accessible training resources within Europe to meet this legal need for knowledge.

Want to know more about:

- Listen to this podcast: https://notebooklm.google.com/notebook/5a1eaf22-860b-4c40-8c78-0a62c71684d5
- Watch this video: https://notebooklm.google.com/notebook/5a1eaf22-860b-4c40-8c78-0a62c71684d5
- Read the full EU AI Act: https://artificialintelligenceact.eu/



What is generative AI (and what is it not)?



A first, simple definition

Generative artificial intelligence (generative AI) refers to digital systems that can generate new content based on patterns learned from very large amounts of data. This content can take different forms: written text, images, audio, video, or combinations of these. In the context of this guidebook, we mainly focus on text-based generative AI,

such as ChatGPT, because of its relevance for lesson planning, language teaching and pedagogical reflection.

What makes generative AI different from earlier digital tools is that it does not simply retrieve existing information or follow fixed rules. Instead, it produces or generates new output each time, based on probabilities: word by word, it predicts what is most likely to come next, given the input it receives. This means that every response is unique and context dependent.

For adult second language teachers, it may be helpful to think of generative AI not as a database or a digital textbook, but as a language-based assistant that can help you think, structure, reformulate, and reflect.

What generative AI is

To use generative AI effectively and confidently, it is important to have a realistic understanding of what it *can* do. In an educational context, generative AI can be described as:

- A text generator that produces language based on patterns learned from massive amounts of written material;
- A support tool that can help teachers draft, structure, adapt or improve lesson plans and teaching materials;
- A thinking partner that can ask clarifying questions, suggest alternatives, or offer different pedagogical approaches;
- A time-saving assistant, especially for repetitive or structurally demanding tasks such as writing lesson objectives, differentiating activities, or adapting language level.

In the GLOW project, teachers experienced that generative AI is particularly useful when it is used to support the thinking and planning process, not to replace it. The quality of the output depends strongly on the quality of the input: the clearer and more pedagogically grounded the prompt, the more useful the result.

What generative AI is not

At the same time, it is equally important to understand what generative AI is not, especially to avoid unrealistic expectations or pedagogical misuse.
Generative AI is not:

- A thinking or understanding being: it does not understand language, meaning or learning in a human way;

- A pedagogical expert with lived classroom experience;
- A neutral or objective source of truth;
- A tool that automatically produces high-quality lesson plans without teacher input;
- A replacement for professional judgement, didactic expertise or knowledge of your learners.

Although AI output often sounds fluent and confident, this does not guarantee correctness or pedagogical suitability. Generative AI does not know your learners, your classroom dynamics, or your institutional context unless you explicitly provide this information.

A common misconception: “AI knows the answer”

A frequent misunderstanding among teachers new to AI is the idea that ChatGPT or similar tools "know" the correct answer. In reality, generative AI does not retrieve answers from a stored knowledge base. Instead, it generates responses by calculating the most probable continuation of text based on its training.

This has two important implications for second language education:

- First, AI can produce plausible but incorrect information. This is why human verification is always necessary. This we call ‘human-in-the-loop' and will be explained in chapter 5 and 6.
- Second, AI output reflects patterns from existing language use, including biases, simplifications and dominant perspectives. This makes critical use essential, especially when working with culturally and linguistically diverse adult learners.

Generative AI and language learning: a natural match?

Language is at the core of generative AI. Large language models (LLM) such as ChatGPT are trained on enormous collections of texts in many languages. As a result, they are particularly strong in tasks related to:

- Reformulating texts at different language levels (e.g. CEFR A1–C1);
- Generating examples, dialogues or scenarios;
- Explaining language use in simple or structured ways;
- Supporting teachers in designing communicative, task-based activities.

However, this strength can also be misleading. Because AI produces fluent language, it may give the impression of linguistic authority. For adult L2 teachers, it is crucial to

remember that fluency does not equal didactic quality. The teacher remains responsible for selecting, adapting and validating all AI-generated material.

A working metaphor: AI as a junior colleague

Within this guidebook, we propose a practical metaphor that resonated strongly with teachers in the GLOW project: think of generative AI as a junior colleague.

A junior colleague:

- Can help with preparation and ideas;
- Needs clear instructions;
- Makes mistakes;
- Improves with feedback;
- Never replaces the experienced teacher.

Approaching AI in this way helps to reduce fear, avoid overreliance, and position the teacher firmly in the role of expert and decision-maker.

In the next chapter, we will look more closely at how ChatGPT actually works, in accessible, non-technical terms, so that you can better understand why prompting matters so much and why AI behaves the way it does.

How does ChatGPT works?

ChatGPT as a language model, not a search engine

ChatGPT belongs to a group of systems called large language models (LLMs). This means that ChatGPT is designed to work with language: it reads text as input and produces text as output. Unlike a search engine, it does not look up information on the internet when you ask a question. Instead, it generates an answer based on patterns it has learned during training.

A useful comparison for teachers is this: a search engine finds *existing* documents, while ChatGPT creates *new* text. This is why its answers do not appear instantly as a finished block of text but are generated word by word.

Let's decode the word ChatGPT

Chat		You have a natural conversation in which you ask a question (=prompt) in your conversation.
G	Generative	It doesn't just search for a real answer (like Google); it generates a new response word by word. It creates a prediction of a pattern
P	Pre-trained	It has been fed a massive amount of data (books, websites, articles) to learn the structures of language.
T	Transformed text or data	ChatGPT does NOT understand in the human sense, It transforms input (text, audio, image) into output (text, audio, image) based on mathematical weights of numbers.

Training on massive amounts of text

ChatGPT has been trained on extremely large collections of texts from many sources, including books, articles, websites and other publicly available material. During this training process, the system learned how words, phrases and sentences typically follow each other in different contexts.

An often-used explanation from the GLOW trainings is helpful here: if a human were to read everything ChatGPT was trained on, this would take several hundred years of continuous reading. This does not mean that ChatGPT remembers all those texts. Instead, it has learned statistical patterns of language use.

For second language teachers, this explains why ChatGPT is often strong at:

- producing grammatically correct sentences;

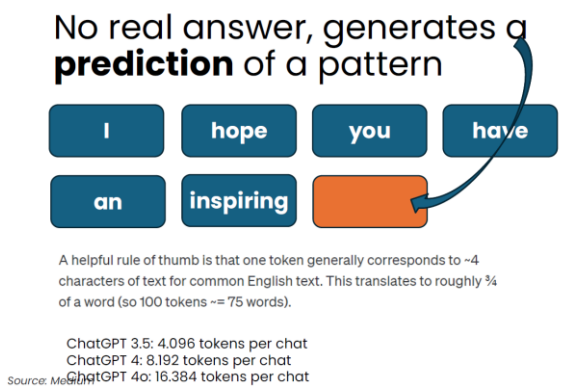
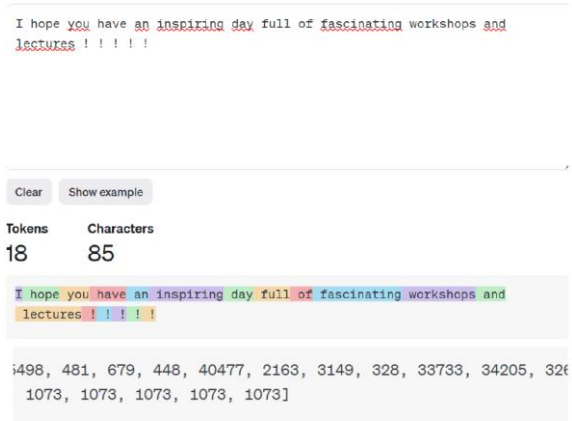
- imitating different registers or tones;
- adapting texts to different CEFR levels;
- generating examples that sound authentic.

From text to numbers – and back again

Although ChatGPT communicates with us through language, internally it does not work with words or meanings. All text is first converted into numbers. These numbers represent so-called tokens, which are pieces of words or symbols.

For example, a short sentence is split into tokens, processed mathematically, and then transformed back into text. This is why ChatGPT does not “understand” language in a human sense. It recognises numerical patterns, not intentions or learning goals.

The Token system

	The AI doesn't read words like we do; it reads "Tokens". Rule of thumb: 1 token ≈ 4 characters (or roughly ¾ of a word). Why this matters: This explains why AI sometimes struggles with tasks like "count the letters in this word" (because it sees the word as one token, not individual letters) or why it has memory limits (context windows are measured in tokens).
	Try it yourself: use the Tokenizer program via this link. The fact that everything is generated can be seen from the output that does not appear in 1 block, but letter by letter.

Prediction, not comprehension

At its core, ChatGPT works by predicting what comes next. Given a prompt, it calculates which word is most likely to follow, then which word follows that one, and so on. This process continues until a complete response is generated.

This explains several classroom-relevant phenomena:

- ChatGPT can give different answers to the same question at different moments;
- Small changes in a prompt can lead to very different outputs;
- The model can sound confident even when it is wrong;
- Asking follow-up questions often improves the result.

What this means for adult second language teachers

Understanding how ChatGPT works leads to a more confident and critical use of the tool. For adult L2 teachers, the key takeaways are:

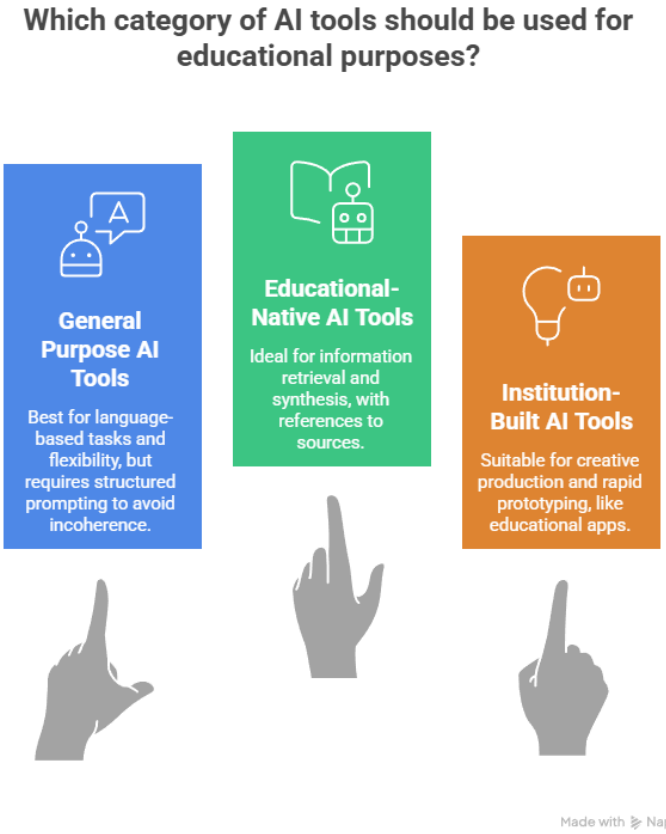
- ChatGPT is strong in language form and structure, not in pedagogy;
- Quality output depends on explicit didactic input from the teacher;
- **AI-generated text always needs human validation and adaptation;**
- Interaction and iteration are more effective than one-off questions.

In the next chapter, we broaden the perspective and look at the main categories of generative AI tools currently used in education, so that teachers can make informed choices depending on their goals.

The big players in generative AI for education

Generative AI is not one single tool, but a rapidly growing ecosystem of applications. For teachers, this abundance can be confusing, especially when many tools appear to overlap in functionality. In this chapter, we group the most relevant tools in the field of education (anno 2025-2026) into three clear categories, based on how they are typically used in educational practice.

This categorisation is intentionally didactic, not technical. It helps teachers understand *what kind of support a tool offers*, rather than how it is built.



Category 1: General purpose AI tools

Chatbots are the most well-known and most widely used generative AI tools in education. They are designed to interact with users through conversation and are particularly suited for language-based tasks.

Typical examples include:

- **ChatGPT** (OpenAI) - available in a free and paid version
- **Gemini** (Google) - available in a free and paid version
- **Claude** (Anthropic) - available in a free and paid version
- **Copilot** (Microsoft) - available in a free and paid version

These tools all share a similar core function: they generate text in response to prompts and follow-up questions. For adult second language teachers, chatbots are especially useful for:

- drafting and structuring lesson plans;
- reformulating texts at different CEFR levels;
- generating examples, dialogues and role plays;

- reflecting on lesson objectives and activities;
- exploring differentiation options.

A key advantage of chatbots is their flexibility. They can adapt to many tasks, provided the teacher gives clear instructions. This is why chatbots played a central role in the GLOW project.

However, this flexibility also comes with a risk: without structure, chatbots may produce output that looks convincing but lacks didactic coherence. This is why structured prompting and rubrics are essential when using these tools for lesson planning.

Category 2: Educational-native AI tools

A second category consists of tools that focus more explicitly on information retrieval, synthesis and explanation, often with references to sources. These tools are less conversational and more research-oriented.

Prominent examples are:

- **NotebookLM** (Google) - free tool
- **Napkin** – free tool
- **Gamma** – free tool

These tools are particularly strong when teachers want to:

- explore a topic and build background knowledge;
- analyse or summarise longer documents;
- work with specific sources (e.g. policy texts, curricula, academic articles);
- prepare explanations or teacher background notes.

Especially NotebookLM can be a very useful co-creator to the teacher because it is very suitable for making direct classroom materials like podcasts, instruction video, flashcards, quizzes, infographics, …

Category 3: Institution-built AI using Vibe-coding and creative AI tools

A third, emerging category consists of tools that focus on creative production and rapid prototyping, sometimes referred to as *vibe-coding*. These tools allow users to create applications like educational apps, interfaces or workflows by describing what they want in natural language.

An example in this category is:

- AI Studio (and similar environments) - free tool
- Custom GPTs – only in the paid version of ChatGPT
- Canva – free tool

These tools are not primarily designed for classroom teaching, but they can be relevant for making educational apps to give to the students to practice their learning.

Within the GLOW context, this category becomes relevant when teachers or organisations start thinking beyond single prompts and move towards custom GPTs or structured AI assistants.

For most adult second language teachers, these tools are not a starting point, but rather a possible future step once confidence with chatbots has been established.

Choosing the right tool: pedagogy first

A central message of this guidebook is that tool choice should always follow pedagogical intent, not technological novelty. No single tool is "best" in all situations.

- General chatbots are most suitable for lesson planning and language-focused tasks;
- Educational-native AI tools support teacher knowledge and preparation;
- Vibe-coding tools enable innovation at organisational level.

For teachers working with adult learners with low digital literacy, simplicity and transparency are crucial. In many cases, using one well-chosen chatbot effectively is pedagogically more powerful than experimenting with many tools superficially.

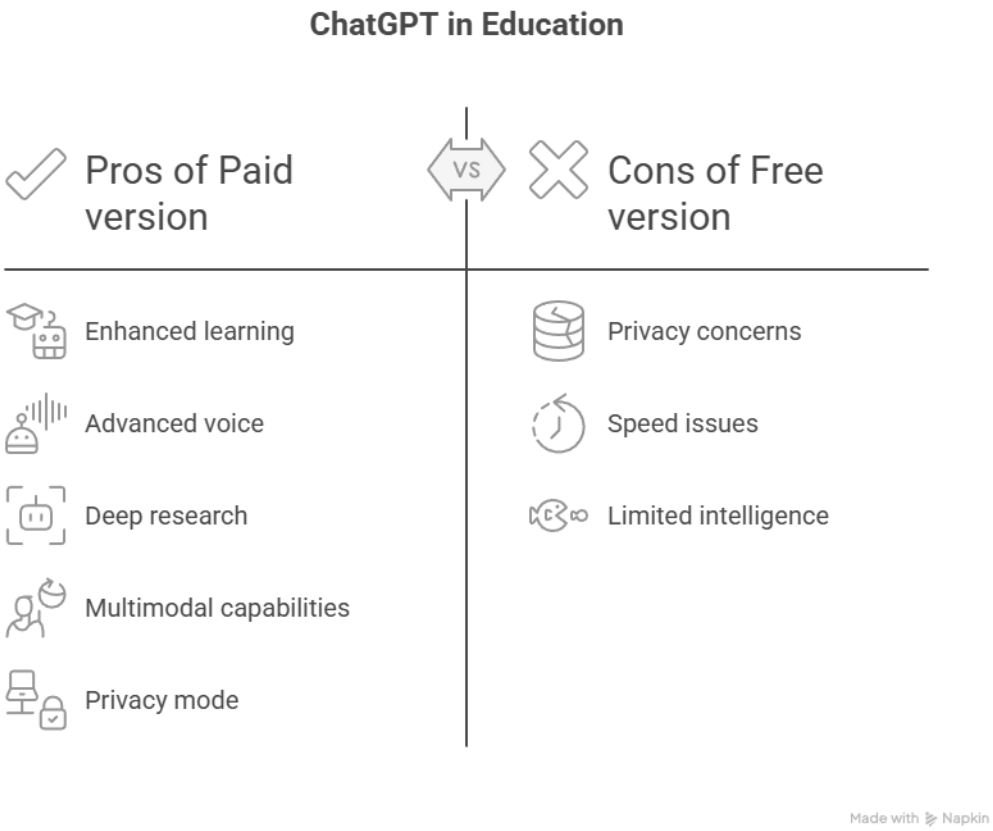
In the following chapters, we therefore focus primarily on chatbots, and in particular on ChatGPT as this in the scope of this project, as a concrete and accessible example for structured, responsible use in adult second language education.

In the scope of the Erasmus+ KA2-project GLOW the focus was on using the chatbot ChatGPT.

Free vs. Paid

Within the scope of the Erasmus+ KA2-project GLOW, one group of teachers was using the free version of ChatGPT and another group was using the paid version of ChatGPT.

Let’s compare the two versions that were available during the project (January 2025 to January 2026).



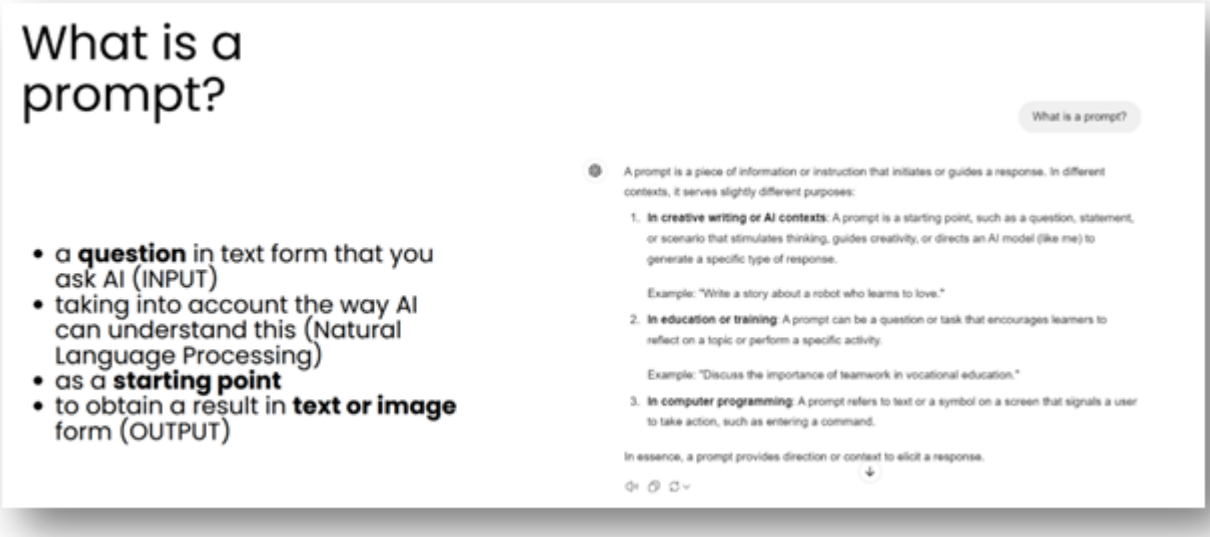
Limitations of Free Versions (Jan 2025 – Jan 2026)

For the "GLOW" teachers using the free version of ChatGPT during this period, the experience is defined by access to capable but "capped" models. While OpenAI upgraded the free tier to include limited access to newer models (like GPT-4o and early versions of GPT-5-mini), significant bottlenecks remain for professional educational use.

From idea to lesson plan: prompting as a pedagogical skill

Why prompting is more than “asking a question”

For teachers who are new to generative AI, prompting is often understood as simply typing a question into a chatbot. Experience from the GLOW project shows that this approach rarely leads to usable lesson plans. Short prompts such as “Make a lesson plan for A2 learners” typically result in generic, superficial output that requires extensive rewriting.



Prompting, as used in this guidebook, should therefore be understood as a pedagogical design activity. When teachers write a prompt, they are not just asking for information; they are translating their didactic intentions into language that an AI system can work with.

In this sense, prompting closely resembles what teachers already do when they:

- formulate learning objectives;
- design lesson sequences;
- anticipate learner needs;
- choose appropriate teaching strategies.

The difference is not pedagogical, but communicative: instead of communicating with learners, the teacher communicates with an AI tool.

Making implicit pedagogy explicit

Much of teachers' professional expertise is implicit. Experienced teachers often "feel" whether an activity will work, without necessarily verbalising all underlying decisions. Generative AI, however, can only work with explicit instructions.

Prompting therefore invites teachers to make their pedagogical reasoning visible:

- Why this task and not another?
- For which learners is it suitable?
- What language level is targeted?
- What kind of support is needed?

This process can initially feel demanding, but many teachers in the GLOW project reported that it led to greater clarity and reflection in their lesson planning, even outside the use of AI.

The 7-step Prompt Engineering Scheme (GLOW model)

Why a structured prompt model is necessary

Within this GLOW-project teachers started to generate their first lesson plans in January 2025 without any training on how to prompt to generate a lesson plan. This was a conscious choice within the project. We wanted to see what the level of prompting was among the teachers involved without testing this beforehand. In February 2025 we delivered the first training on prompt engineering to the teachers involved in the project.

One of the most important insights from the GLOW project is that quality does not come from the tool itself, but from the structure of the prompt. Teachers across all three research groups (no AI, free AI, paid AI) observed that unstructured prompts led to inconsistent results, regardless of the version used.

The 7-step Prompt Engineering Scheme was therefore developed to help teachers:

- think pedagogically before prompting;
- communicate clearly with the AI tool;
- obtain more consistent and reusable output;
- reduce time spent rewriting prompts.

Importantly, this model does not require technical knowledge. It builds directly on didactic reasoning that teachers already apply when designing lessons.

Overview of the GLOW 7-step model

The GLOW model consists of seven clearly defined elements developed by Lut De Jaegher from Artevelde – University of Applied Science:

1. Role
2. Task
3. Context
4. Target group
5. Parameters
6. Tone of voice
7. Extra information

These steps do not have to be written as separate bullet points in the final prompt, but they should all be present when a teacher formulates a prompt.

In the sections below, we explain each step from the perspective of adult second language education.

Step 1 – Role: who should the AI be?

The first step defines the **role or persona** that the AI should adopt. This helps the system frame its response from a specific perspective.

For example, instead of writing:
"*Make a lesson plan for Dutch A2 learners*",

a teacher can specify:
"***Act as an experienced teacher of Dutch as a second language for adult learners with 20 years of experience in effective pedagogy.***"

Defining the role:

- improves relevance of the output.
- reduces generic responses.
- aligns AI output with professional language and practice.

The role does not describe who *you* are as a teacher, but who the AI should pretend to be.

Step 2 – Task: what should the AI do?

The task clearly states **what kind of output** you expect. Vague tasks lead to vague results.

Examples of well-defined tasks include:

- develop a lesson plan;
- reformulate an existing activity;
- suggest differentiation options;
- review a lesson plan critically.

For adult L2 teachers, it is often helpful to specify the mandatory components of the task (e.g. lesson objectives, warm-up, main activities).

Step 3 – Context: in which situation will this be used?

Contextual information situates the task in a **real teaching situation**. This includes elements such as:

- type of course (integration, workplace language, general L2);
- teaching format (face-to-face, online, hybrid);
- lesson duration;
- available infrastructure.

Context helps the AI avoid unrealistic suggestions and increases practical usability.

Step 4 – Target group: who are the learners?

Adult second language classrooms are rarely homogeneous. The target group step allows teachers to describe learner characteristics such as:

- CEFR level;
- age range;
- literacy level;
- learning needs or constraints.

This step is crucial for differentiation and inclusion. Even a short description can significantly improve output quality.

Step 5 – Parameters: what constraints apply?

Parameters define **boundaries and formats**. They prevent AI output from becoming too long, too complex or pedagogically unfocused.

Examples include:

- lesson duration;
- number of activities;
- inclusion of listening or speaking tasks;
- use of specific frameworks (CEFR, Bloom's taxonomy).

Teachers in the GLOW project reported that clear parameters were one of the strongest predictors of usable output.

Step 6 – Tone of voice: how should it sound?

Tone of voice determines **style and register**. In lesson planning, this often concerns whether output should be:

- formal or informal;
- concise or elaborated;
- instructive or reflective.

Specifying tone helps ensure that AI output fits the teacher's professional context and personal style.

Step 7 – Extra information: interaction and refinement

The final step invites interaction. Teachers can ask the AI to:

- ask clarifying questions;
- wait for feedback before finalising output;
- revise its own response.

This step reinforces the **human-in-the-loop principle** and turns prompting into a dialogue rather than a one-off request.

The 7-step prompt engineering scheme is explaining during the first online training in the GLOW-project. You can watch this training using this link.

Comprehensive rubric for evaluating lesson plan prompts

The partners of the KA2 GLOW-project discussed what elements are mandatory in a lesson plan and which are nice-to-have.

The mandatory elements are:

- **General Information** because this sets the scope for the lesson and aligns expectations for learners and teachers. It helps to ensure the lesson meets the learners' specific needs (e.g., beginner level adults preparing for practical scenarios).
- **Lesson Objectives** because they ensure focus and enable measurable progress. Clear, measurable objectives help both the teacher and learners understand the desired outcomes. It focuses on the lesson, provides direction, and allows for effective assessment at the end.
- **Warm-up Activity** because it activates prior knowledge, helps learners feel comfortable, and sets the context for the lesson. It also promotes peer interaction and lowers affective filters (anxiety in language learning).
- **Body of the lesson** divided in different activities: Introduces vocabulary and phrases in context.

The nice-to-have elements are:

- Materials
- Assessment
- Reflection/Homework

To come to a qualitative lesson plan, a good prompt in ChatGPT can help teachers to reach this goal. That's why the rubric below was developed in the framework of the KA2 GLOW-project.

This rubric evaluates the quality of prompts used to generate AI-assisted lesson plans. It integrates key aspects of specificity, complexity, and effectiveness while incorporating elements such as audience targeting, taxonomic depth, active learning strategies, and contextual considerations. The scale ranges from 0 (Poor) to 3 (Excellent). The rubric used during the project can be found in the appendices.

Examples of basic and advanced prompts

After delivering the first training in February 2025 on prompt engineering to the teachers involved in the GLOW project we noticed some things. First, we saw that many teachers using either the free or paid version of ChatGPT formulated their prompts in one or two

sentences and then expected a fully developed lesson plan. Second, we saw that teachers copied and pasted the generated output, which didn't immediately align perfectly with their needs or the sample template used, into the suggested template to make it fit within that template. This resulted in unnecessary, pointless extra work.

Finally, we saw that teachers using only the free version often failed to develop the desired lesson plan due to the limited time available for prompting. Due to the time pressure to provide extensive prompting, they are not always able to adjust the prompt step by step. As a result, they are less likely to produce a high-quality lesson plan that is focused on content and sufficiently adapted to learning objectives, the target group, and the available lesson time.

Therefore, it was decided to organize a second training to teach teachers how to achieve the desired result based on an advanced prompt instead of a simple basic prompt.

Example of a prompt that we will improve step by step

> From now on you are a teacher of Dutch as a second language. Tomorrow you will give a first lesson to a level B1 Threshold 4 oral. These are students who have a strong foundation in Dutch. I want you to prepare a lesson for these students. There is a digital board with audio and one computer in the classroom. The aim of the lesson is for the students to get to know each other. They repeat the imperfect and the perfect tense in terms of grammar (they already learned this in the previous level). The theme of the lesson is annoyances in daily life. The lesson lasts 180 minutes. The students must learn how to correctly formulate annoyances on the street orally. There should also be a number of listening fragments. Can you provide this in a downloadable Word document?

Basic prompt

1. Role
2. Task
3. Context
4. Targetgroup
5. [redacted]
6. Tone of voice
7. Extra info

From now on you are a teacher of Dutch as a second language. Tomorrow you will give a first lesson to a level B1 Threshold 4 oral. These are students who have a strong foundation in Dutch. I want you to prepare a lesson for these students. There is a digital board with audio and one computer in the classroom. The aim of the lesson is for the students to get to know each other. They repeat the imperfect and the perfect tense in terms of grammar (they already learned this in the previous level). The theme of the lesson is annoyances in daily life. The lesson lasts 180 minutes. The students must learn how to correctly formulate annoyances on the street orally. There should also be a number of listening fragments. Can you provide this in a downloadable Word document?

Improved prompt

1. Role
2. Task
3. Context
4. Targetgroup
5. Parameters
6. Tone of voice
7. Extra info

#**Role**: You are a teacher of Dutch as a second language with more than 20 years of experience and an expert in developing lesson plans with high quality and impact on effective learning and teaching also with more than 20 years of experience.

#**Task**: Develop a qualitative lesson plan about the theme 'annoyances in daily live' consisting the following elements mandatory elements:

- **General Information** (lesson title, target group, proficiency level, duration, class size and F2F, online of hybrid context).
- **Lesson Objectives** (SMART formulated)
- **Warm-up Activity**
- **Body of the lesson divided in different activities**

The nice-to-have elements in the lesson plan are:

- **Materials**
- **Assessment**
- **Reflection/Homework**

#**Targetgroup**: The group of students consist of 20 adult students between 18 and 35 years old, following level B1 Threshold 4 oral and have already a strong foundation in Dutch.

#**Context**:

Tomorrow will be the first lesson of the module B1 Threshold 4 oral. The aim of the lesson is to get to know each other. They need to repeat the imperfect and the perfect tense. They already learned this in the previous level. They must learn how to correctly formulate annoyances on the street orally.

- Specificity of learning objectives: Formulate the 3 learning objectives SMART with concrete examples and explicit connected to the subject-specific domains with clear progression indicators.
- Taxonomic depth: Reference to Bloom's taxonomy to have detailed cognitive progression. Give clear level indicators and integrate with the subject-specific skills.
- Active learning strategies: Suggest 3 active learning strategies that works well for the theme of this lesson. Also explain why you suggest these active learning strategies. Make sure that differentiation on the cognitive level of the students is possible and explain also why you suggest this differentiation.

#**Parameters**: Forsee in the lesson plan 3 listening fragments. The total duration of the lesson is 180 minutes. There is a digital board with audio and one computer in the classroom. Students can use their smartphone.

#**Tone**: Implicitly formal and instructive, teacher-to-AI tone

#**Extra info**:

Set up the lesson plan first in the Canvas so that I can adjust it. Wait to make the final lesson plan until I say it.

Finally give me the whole lesson plan in a downloadable word-document starting with the prompt that was used to generated the lesson plan.

The second training session can be watched on the website.

Chapter 3. What are teachers telling us

Teaching Experience

Teachers participating in the surveys demonstrate extensive professional experience. Responses from KOV teachers indicate an average teaching experience of approximately 15 years, with individual experience ranging from roughly two years to more than twenty years. This range suggests a workforce composed of both early-career teachers and long-established professionals.

Similarly, Briga teachers reported a wide spectrum of teaching experience. The presence of both newer teachers and highly experienced educators creates an environment in which pedagogical innovation and accumulated professional knowledge coexist. Experienced teachers often rely on well-established teaching strategies and long-term observations of student learning patterns. Meanwhile, less experienced teachers frequently report experimenting with new digital tools and instructional approaches.

Experience with Digital Tools

Across both organisations, teachers report significant experience using digital technologies in their teaching practice. In the KOV dataset, teachers reported an average of approximately nine years of experience working with digital tools. This indicates that digital technologies have been integrated into second language education for a considerable period of time.

Briga teachers similarly report long-term familiarity with digital platforms, learning management systems, and interactive applications. Teachers describe using digital tools for a variety of purposes, including delivering lessons online, creating exercises, communicating with students, and supporting independent practice.

However, despite this high level of experience, teachers frequently highlight the rapid pace of technological change as a challenge. Many respondents emphasize the difficulty of staying up to date with new platforms, updates, and digital teaching practices.

Levels Taught in Second Language Learning

Teachers across both organisations work with learners across a wide range of language proficiency levels. Most respondents teach levels ranging from A1 (beginner) to B2 (upper-intermediate), although some teachers also report experience with more advanced learners.

Teaching across multiple levels requires considerable pedagogical flexibility. Beginner learners require structured instruction, repetition, and visual support, while

intermediate and advanced learners require opportunities for discussion, abstract thinking, and extended language production. Teachers report that managing such diversity within the same institution often requires the adaptation of teaching materials and strategies.

Cognitive Differences in Adult Second Language Learning

Impact of Cognitive Differences on Learning Pace

Teachers consistently report that cognitive differences strongly influence the pace of language acquisition. In the KOV responses, the majority of teachers indicated that cognitive differences affect learning pace “very often” (rating 4) or “always” (rating 5). Only a small minority indicated that these differences occur only occasionally.

This finding highlights the reality that adult language classrooms often consist of highly heterogeneous groups of learners. Differences in memory capacity, attention span, prior educational experience, and learning strategies can create large disparities in learning speed.

Assessing Cognitive Strengths and Weaknesses

Teachers report using a combination of formal and informal assessment strategies. Classroom observation remains the most common approach. Teachers monitor how learners respond to instructions, how they participate in activities, and how they perform during speaking or writing tasks.

Additional methods include diagnostic tests, formative quizzes, written assignments, and reflective exercises. Teachers also analyze patterns of errors to identify areas where learners struggle.

Differentiated Instruction

The majority of teachers in both organisations report using differentiated instruction. Common differentiation strategies include providing tasks with varying difficulty levels, adjusting the pace of instruction, and offering optional extension activities for advanced learners.

Despite its perceived effectiveness, teachers frequently describe differentiated instruction as time-intensive. Preparing multiple versions of exercises and providing individual feedback requires significant preparation time.

Challenges in Adapting Teaching Methods

Teachers identify several key challenges when adapting teaching methods to diverse cognitive abilities:

- Limited preparation time
- Large and heterogeneous class groups
- Differences in digital literacy
- Lack of appropriate digital resources
- Managing the pace of lessons for both stronger and weaker learners

Many teachers emphasize that weaker learners often require additional repetition and support, which can slow down the overall pace of the class.

Support Strategies for Slower Learners

Teachers employ a variety of strategies to support slower learners, including:

- Step-by-step instruction
- Frequent repetition of key concepts
- Visual aids such as images and diagrams
- Peer support systems
- Simplified instructions and tasks

Confidence building is also considered crucial. Teachers emphasize creating a safe learning environment where mistakes are accepted as part of the learning process.

Peer Learning and Group Work

Peer learning is generally perceived as effective, although its success depends on group composition and task structure. Teachers report that stronger learners can help explain concepts to weaker learners, which benefits both groups.

However, when cognitive gaps between learners are very large, peer learning can become less effective.

Technological Tools

Teachers report using a wide range of technological tools to support differentiated learning. Examples mentioned include:

- Learning management systems such as Canvas
- Interactive exercise tools like BookWidgets
- Vocabulary tools such as Quizlet
- Online quiz platforms like Kahoot
- Video platforms including YouTube
- Collaborative tools such as Padlet

Some teachers emphasize the importance of limiting the number of digital tools used in a course in order to avoid cognitive overload.

Common Cognitive Challenges

Teachers identify several common challenges faced by adult learners with limited educational backgrounds:

- Difficulty understanding grammar concepts
- Limited working memory and concentration
- Lack of effective study strategies
- Limited digital literacy
- Stress related to work, migration, or family responsibilities

These challenges highlight the importance of structured teaching methods and realistic learning expectations.

Knowledge and Opinions about Generative AI

Familiarity with Generative AI

Teachers' familiarity with generative AI varies significantly. Some teachers report being very familiar with AI tools such as ChatGPT or Microsoft Copilot, while others report only minimal knowledge.

In the KOV responses, participants rated their familiarity across a spectrum ranging from "not familiar at all" (rating 1) to "very familiar" (rating 4). A substantial group of teachers described themselves as only slightly familiar with generative AI.

Use of Generative AI in Teaching

Several teachers report using generative AI tools, primarily the free versions of platforms such as ChatGPT, Copilot, Claude, and Gemini. These tools are typically used for:

- Generating exercises
- Simplifying texts
- Brainstorming lesson ideas
- Creating example sentences
- Adapting reading materials

However, a significant number of teachers report that they have not yet used generative AI in their teaching.

Confidence in Using AI

Confidence levels vary widely. Teachers who have experimented with AI tools tend to report higher confidence, while those with limited exposure often express uncertainty.

Several teachers note that learning how to write effective prompts requires practice and training.

Perceived Challenges

Teachers identify several concerns regarding generative AI:

- Reliability of AI-generated content
- Risk of students becoming overly dependent on AI
- Ethical concerns regarding academic integrity
- Loss of creativity in teaching materials
- Environmental concerns related to AI technologies

Teachers also emphasize that AI cannot replace the observational and emotional aspects of teaching.

Potential for Workload Reduction

Opinions are mixed regarding whether AI reduces workload. Some teachers report that AI helps them generate exercises quickly and provides inspiration for lesson planning. Others note that verifying and refining AI outputs can require additional time.

Training and Professional Development

Training opportunities appear uneven. Some teachers report attending workshops focused on prompts, AI tools, or ethical considerations. However, many respondents indicate that they have received little or no formal training.

Teachers consistently express interest in practical training sessions that focus on real classroom applications.

Overall summary of feedback

Several key patterns emerge from the combined analysis of feedback from Briga and KOV teachers.

First, cognitive diversity is a defining characteristic of adult second language classrooms. Teachers consistently report that differences in prior education, cognitive abilities, and life circumstances require constant adaptation of teaching strategies.

Second, while digital tools are already widely used in language education, teachers continue to face challenges related to technological change, resource availability, and digital literacy among learners.

Third, generative AI is currently in an exploratory phase within second language teaching. Some teachers actively experiment with AI tools and see clear potential

benefits. Others remain cautious due to concerns about reliability, ethics, and the potential impact on student learning.

Finally, the data highlights the importance of professional development. Teachers express strong interest in practical training that helps them integrate digital tools and AI technologies effectively into their teaching practice.

Overall, the findings suggest that AI tools designed specifically for second language education—particularly those capable of supporting differentiated instruction—could play a valuable role in addressing many of the challenges identified by teachers.

Chapter 4. Results of Integrating AI into Lesson Planning

GLOW Project Results - Belgium

Prepared for the GLOW / GAHWP-LT project. **Sources used:** KA210 project description, baseline KOV teacher synthesis, Belgian longitudinal Excel dataset, and the earlier Belgium results document for hypothesis framing. This version applies the same cleaning logic as the Norway report to improve cross-country comparability.

167	17	11
observations	teachers	project months

Executive summary

- This report analyses the Belgian dataset with the same harmonised cleaning logic used for Norway. One summary row was removed, and unrated outputs with blank rubric dimensions were treated as missing rather than as true zero scores.
- The Belgian strand contains a full three-group design: Control, Free AI and Paid AI. This makes it possible to test both AI-versus-control and paid-versus-free hypotheses.
- Paid AI produced the strongest lesson-plan quality descriptively (M = 10.76/12), but that paid-versus-free advantage was not statistically reliable once repeated measures and time were modelled.
- Prompt quality rose most clearly after the advanced prompt training around month 5 and then remained high. The month-10 custom GPT session appears to have stabilised performance rather than creating another large jump.
- The strongest Belgian effects are longitudinal: AI users improved lesson-plan quality over time and also became faster. The broader AI-versus-control advantages on quality, structure and engagement remain limited.

Belgium results

Project context and analytical focus

The GLOW / GAHWP-LT project aims to reduce workload and stress for adult language teachers by improving the way generative AI is used for lesson planning. In the Belgian strand, the project design is methodologically stronger than in Norway because it includes three conditions: a control group without ChatGPT, a Free AI group and a Paid AI group. This permits both AI-versus-control comparisons and free-versus-paid comparisons, alongside longitudinal learning effects across the project months.

The baseline synthesis “What are KOVs teachers telling us” shows that Belgian teachers entered the project with substantial professional experience: around 15 years of teaching on average and around 9 years of work with digital tools, across language levels from A1 to B2 and beyond. At the same time, familiarity with generative AI was uneven. Some teachers had already experimented with free tools such as ChatGPT or Copilot, whereas others had not used AI at all. Baseline concerns centred on reliability, student over-reliance, loss of critical thinking and the need for practical, classroom-oriented training.

That starting point makes the three training moments analytically central: month 2 introduced basic prompt writing, month 5 focused on richer and higher-quality prompts, and month 10 introduced the use of a custom GPT for lesson-plan building. The present report therefore gives special attention to phase-based prompt development, the evolution of lesson-plan quality, and the extent to which AI use translated into time savings.

Method

Dataset overview

After harmonised cleaning, the Belgian analytical file contained 167 observational rows across 21 variables, collected from 17 teachers over 11 project months with data (months 1–7 and 9–12). One spreadsheet summary row (coded as month 167) was removed because it did not represent a real observation. The cleaned file contains 57 control observations, 72 Free AI observations and 38 Paid AI observations.

To align the Belgian analysis with the Norway report, outputs were treated as missing when a total score of 0 was recorded but all rubric dimensions were blank. This affected five lesson-plan totals and five prompt totals, which were interpreted as unrated submissions rather than true zero-quality products. One teacher switched from Free AI to Paid AI after month 2; group assignment was therefore retained at the row level so that each lesson plan was analysed under the actual tool condition recorded for that month.

Indicator	Value
Analytical observations	167
Teachers	17
Project months with data	1–7 and 9–12
Groups recorded	Control (6 teachers), Free AI (8 teachers), Paid AI (4 teachers)
Crossover participants	1 teacher (kov-25) moved from Free AI to Paid AI after month 2
Non-observational rows removed	1 summary row (month 167)

Missing data analysis

Missingness was low for the main evaluative outcomes. The preparation-time variable showed the highest non-structural missingness (n = 9, 5.39%), and all nine missing time

observations occurred in the control group. The four lesson-plan rubric dimensions and the final lesson-plan total each had five missing values (2.99%), corresponding to unrated submissions. Prompt missingness was much higher in percentage terms (n = 62, 37.13%), but almost all of this was structural because control-group teachers did not produce prompts; the remaining five missing prompt totals were blank early Free AI submissions.

Because the Belgian data are longitudinal and unbalanced, all inferential analyses used all available observations with teacher-level repeated-measures models. For most hypotheses, linear mixed-effects models with random intercepts per teacher were used. For the phase-based prompt analysis, GEE estimates are additionally reported because they provide a stable clustered estimate for the phase comparisons. Preparation time was strongly right-skewed, so time hypotheses were checked both in raw minutes and on the log-transformed scale.

Variable	Missing n	Missing %
Time spent on making the lesson plan	9	5.39%
Structure & Clarity	5	2.99%
Alignment with (Curriculum) Standards	5	2.99%
Engagement Strategies	5	2.99%
Clarity & Completeness	5	2.99%
Final score lessonplan	5	2.99%
Final score prompt	62	37.13%

Researchable hypotheses

Unlike Norway, the Belgian dataset allows all of the main project hypotheses to be tested directly. The present report therefore evaluates the original Belgium hypotheses H1, H2, H4, H5, H6, H7, H8, H9, H10, H11 and H14, while keeping the interpretation aligned with the Norway report so that a future cross-country comparison can be made on a more comparable basis.

Results

Descriptive statistics

Across rated Belgian lesson plans, the mean lesson-plan quality score was 9.48 (SD = 2.66, N = 162) on the 0–12 lesson-plan rubric, equivalent to 79.0% of the maximum possible score. The median was 10 and observed values ranged from 2 to 12. Across AI observations with rated prompts, the mean prompt score was 18.88 (SD = 4.84, N = 105) on the 0–24 prompt rubric. Preparation time averaged 72.1 minutes (SD = 61.4, N = 158), with a median of 57.5 minutes and a wide range from 5 to 400 minutes, confirming a strongly right-skewed time distribution.

Group-level descriptive statistics show the clearest quality advantage for the Paid AI condition. Paid AI produced the highest lesson-plan quality (M = 10.76; 89.7% of the maximum rubric score), followed by Free AI (M = 9.16) and Control (M = 8.99). The paid group also obtained slightly stronger prompt scores than the free group (M = 19.51 versus 18.51). For preparation time, however, the pattern reversed: Free AI was by far the fastest condition on average (M = 51.3 minutes), whereas Paid AI remained much slower (M = 85.9 minutes) and closer to the Control group (M = 92.5 minutes). This suggests that premium access did not automatically translate into a more efficient workflow.

Outcome	Scale	N	Mean	SD	Median	Range
Prompt score	0-24	105	18.88	4.84	20.0	4.0-24.0
Lesson plan quality	0-12	162	9.48	2.66	10.0	2.0-12.0
Lesson plan quality (% max)	0-100%	162	78.99	22.19	83.3	16.7-100.0
Preparation time (minutes)	Raw minutes	158	72.14	61.36	57.5	5.0-400.0

Group	Teachers	Obs.	Prompt mean	Prompt SD	Lesson mean	Lesson SD	Lesson % max	Time mean	Time SD	Rated lessons	Time N
Control	6	57			8.99	2.16	74.92	92.51	45.33	54	48
Free AI	8	72	18.51	4.65	9.16	3.17	76.31	51.28	48.18	70	72
Paid AI	4	38	19.51	5.18	10.76	1.78	89.69	85.92	85.64	38	38

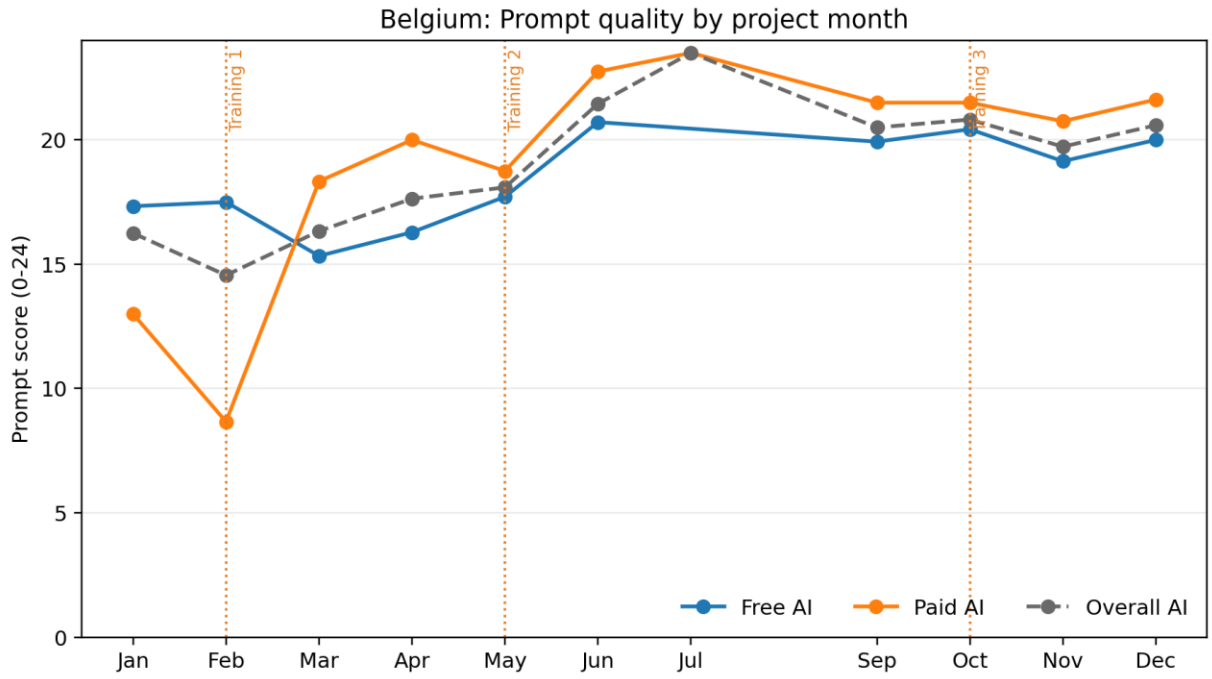


Figure 1. Monthly mean prompt quality in Belgium, with training moments marked at months 2, 5 and 10. The prompt figure includes only AI observations. July and August were not expected work months.

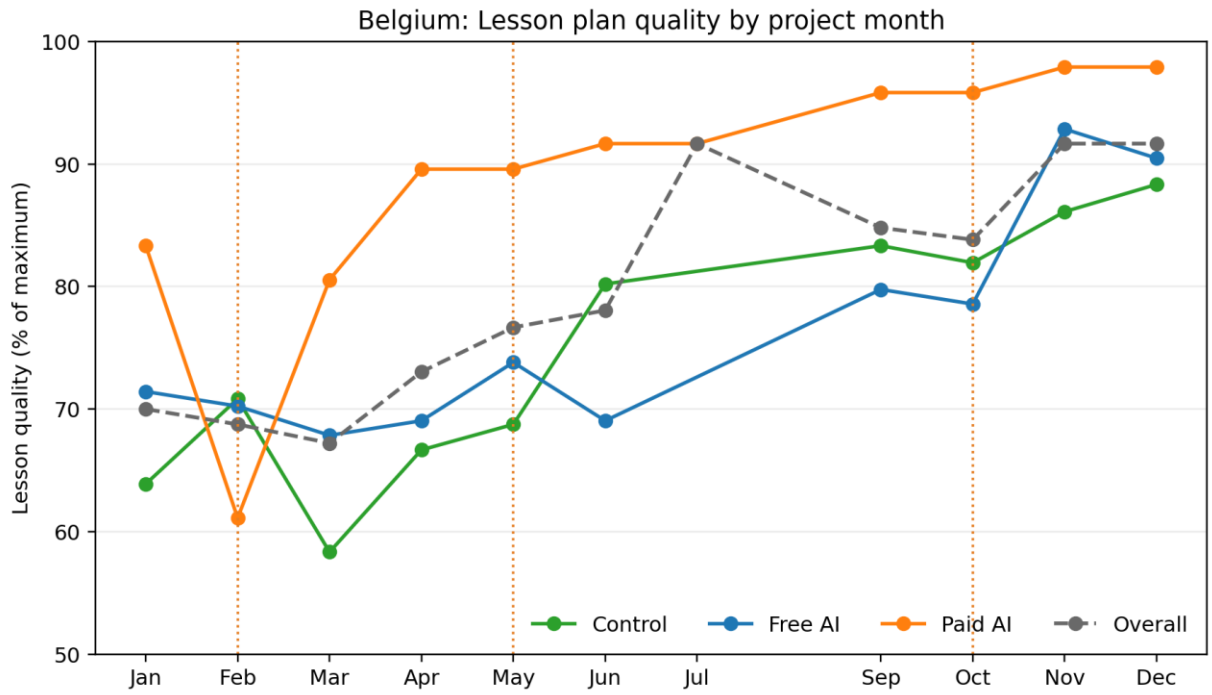


Figure 2. Monthly mean lesson-plan quality in Belgium, expressed as a percentage of the Belgian maximum score. July should be read cautiously because it contains only two Paid AI observations. July and August were not expected work months.

Training-phase analysis

To reflect the staged training design, Belgian AI observations were divided into four phases: Phase 0 (month 1, before training), Phase 1 (months 2–4, after the basic prompt session), Phase 2 (months 5–9, after the advanced prompt session), and Phase 3 (months 10–12, after the custom GPT session). The phase means show a clear pattern. Prompt quality was essentially flat between Phase 0 and Phase 1 (16.25 to 16.28), then increased strongly in Phase 2 (20.21) and remained high in Phase 3 (20.38). Lesson-plan quality improved across phases from 8.89 to 10.91, while average preparation time fell from 81.0 to 48.1 minutes.

The inferential phase model confirms that the first training did not produce a clear step change, but the later stages did. Relative to Phase 0, the GEE phase model estimated no meaningful gain for Phase 1 (+0.33, p = .842), but significant gains for Phase 2 (+4.34, p = .022) and Phase 3 (+4.54, p = .026). The contrast between Phase 3 and Phase 2 was negligible (+0.21, p = .808), indicating that the month-10 custom GPT session stabilised a higher level rather than generating another sharp jump.

Phase	Description	Prompt N	Prompt mean	Lesson N	Lesson mean	Time N	Time mean
Phase 0	Before training	8	16.25	9	8.89	10	81.0
Phase 1	After training 1 (months 2-4)	29	16.28	31	8.65	32	75.84
Phase 2	After training 2 (months 5-9)	35	20.21	35	9.77	35	60.93

Phase 3	After training 3 (months 10-12)	33	20.38	33	10.91	33	48.12

Results of the hypotheses

H1: Lesson plans created with AI (free or paid) have a higher quality score on average than lesson plans without AI.

Descriptively, AI-supported lesson plans scored slightly higher than non-AI lesson plans (M = 9.72 versus 8.99). In a simple unadjusted comparison, this difference was small and only borderline (Welch p = .072, d = 0.28). In the repeated-measures mixed model controlling for month, the AI coefficient was +0.58 points and non-significant (p = .505). H1 is therefore not supported. In the Belgian strand, having AI available did not by itself guarantee substantially better lesson-plan quality.

H2: Lesson plans created with a paid AI tool have higher quality than lesson plans created with a free AI tool.

Paid AI outperformed Free AI descriptively (M = 10.76 versus 9.16), and the simple between-group comparison was significant (Welch p = .001, d = 0.58). However, in the repeated-measures mixed model controlling for month, the paid coefficient (+0.69) was not significant (p = .485). H2 therefore receives partial support. The paid condition looks stronger descriptively, but the advantage is not statistically reliable once time and within-teacher clustering are taken into account.

H4: Prompt quality increases after each training (months 2, 5 and 10).

The harmonised Belgian re-analysis gives partial support to H4. Phase 1 did not differ meaningfully from the pre-training baseline (+0.33, p = .842), whereas Phase 2 (+4.34, p = .022) and Phase 3 (+4.54, p = .026) were both significantly higher than baseline. The planned contrast between Phase 3 and Phase 2 was negligible (+0.21, p = .808). In practice, this means that the advanced prompt training around month 5 mattered more than the initial training, and the month-10 custom GPT session consolidated the gain rather than producing a new jump.

H5: Teachers who use AI spend less time creating a lesson plan than teachers without AI.

AI users were descriptively faster than control teachers (63.3 versus 92.5 minutes; simple Welch p = .0015). In the repeated-measures models, however, the result depended on the scale used for time: in raw minutes the AI coefficient was -30.75 and non-significant (p = .219), but on log-transformed time it was -0.62 and just significant (p = .049). H5 therefore receives mixed or weak support rather than a clean confirmation. The practical pattern is compatible with time savings, but the statistical evidence is not fully robust, especially because all nine missing time values occur in the control group.

H6: Teachers who use a paid AI tool spend less time than teachers who use a free AI tool.

The descriptive pattern ran in the opposite direction to the hypothesis. Paid AI users spent more time than Free AI users on average (85.9 versus 51.3 minutes; simple p = .025 in the opposite direction). In adjusted mixed models, however, the paid-versus-free time difference was not significant on either the raw-minute or log-time scale. H6 is not supported. Paying for AI did not make Belgian teachers faster.

H7: The quality of lesson plans increases over time among AI users.

The repeated-measures mixed model showed a clear positive month effect on lesson-plan quality among AI users (b = +0.262 points per month, p < .001). This aligns with the descriptive pattern from around 8.9 points in month 1 to above 11 points by months 11 and 12. H7 is supported. The Belgian data show a robust learning effect in lesson-plan quality across the project period.

H8: The time needed to create a lesson plan decreases over time among AI users.

The time models both point in the same direction. The raw-time model estimated a decrease of 3.55 minutes per month (p < .001), and the log-time model confirmed a significant negative slope (b = -0.051, p < .001). H8 is supported. Belgian teachers became progressively more efficient in using AI-supported lesson planning over time.

H9: Higher prompt quality leads to higher lesson-plan quality.

At the descriptive level, prompt quality and lesson-plan quality were positively correlated in Belgium (r = .329, p < .001). In the adjusted mixed model, however, the prompt coefficient was +0.090 and only marginal (p = .085) after controlling for month and the paid/free distinction. H9 therefore receives weak or descriptive-only support in this harmonised re-analysis. Better prompts tend to coincide with better lesson plans, but the independent prompt effect is more modest than the earlier Belgium report suggested.

H10: Using AI leads to better structure and clarity in lesson plans.

The adjusted model for the Structure & Clarity rubric dimension showed no AI advantage. The AI coefficient was -0.133 (p = .622), while the month effect was positive, indicating that both AI and control teachers improved structurally over time. H10 is not supported.

H11: AI usage increases the use of engagement strategies.

The adjusted model for Engagement Strategies also showed no reliable AI effect. The AI coefficient was +0.192 and non-significant (p = .527). H11 is not supported. Belgian lesson plans became more engaging over time, but not because AI users pulled away from the control group.

H14: The difference between AI and non-AI groups increases as the project progresses.

The AI × month interaction for lesson-plan quality was essentially zero (-0.012, p = .885). Both AI and non-AI teachers improved over time, but the gap between them did not widen systematically. H14 is not supported.

Overview of all hypotheses

Hypothesis	Statement	Testable in Belgium?	Result	Comment
H1	AI-supported lesson plans score higher than non-AI lesson plans	Yes	No	Descriptively slightly higher for AI, but adjusted group effect not significant
H2	Paid AI lesson plans score higher than free AI lesson plans	Yes	Partial support	Large descriptive advantage for paid AI; adjusted repeated-measures model not significant
H4	Prompt quality increases after trainings in months 2, 5 and 10	Yes	Partial support	No clear gain after training 1; clear gains by phases after training 2; no added gain after training 3
H5	AI users spend less time than non-AI teachers	Yes	Mixed / weak support	Descriptively faster and significant on log-time model, but not significant on raw-minute model
H6	Paid AI users spend less time than free AI users	Yes	No	Descriptive pattern runs opposite to the hypothesis; adjusted models not significant
H7	Lesson plan quality improves over time among AI users	Yes	Yes	Positive month effect in mixed model
H8	Time spent decreases over time among AI users	Yes	Yes	Negative month effect in both raw-time and log-time mixed models
H9	Higher prompt quality leads to higher lesson plan quality	Yes	Weak / descriptive only	Positive correlation, but adjusted mixed model only marginal (p=.085)
H10	AI leads to better structure and clarity	Yes	No	No significant AI-versus-control effect
H11	AI leads to more engagement strategies	Yes	No	No significant AI-versus-control effect
H14	The AI/non-AI difference grows over time	Yes	No	AI × month interaction not significant

Technical appendix: key repeated-measures model outputs

The table below summarises the main inferential models used in the Belgian re-analysis. Linear mixed-effects models with random intercepts per teacher were used for most hypotheses. For H4, GEE phase estimates are reported because they provide a stable clustered estimate for the phase comparisons.

Hypothesis	Model	Parameter	Estimate	p-value
H1	Mixed model: lesson score ~ AI use + month	AI coefficient	0.585	0.505
H2	Mixed model: lesson score ~ paid AI + month	Paid AI coefficient	0.686	0.485
H4	GEE: prompt score ~ training phase	Phase 1 vs baseline	0.334	0.842
H4	GEE: prompt score ~ training phase	Phase 2 vs baseline	4.335	0.022
H4	GEE: prompt score ~ training phase	Phase 3 vs baseline	4.544	0.026
H4	GEE planned contrast	Phase 3 vs Phase 2	0.209	0.808
H5	Mixed model: log time ~ AI use + month	AI coefficient	-0.619	0.049
H6	Mixed model: log time ~ paid AI + month	Paid AI coefficient	0.045	0.865
H7	Mixed model: lesson score ~ month	Month slope	0.262	<0.001
H8	Mixed model: log time ~ month	Month slope	-0.05	<0.001
H9	Mixed model: lesson score ~ prompt score + month + paid AI	Prompt score coefficient	0.09	0.085
H10	Mixed model: structure score ~ AI use + month	AI coefficient	-0.133	0.622
H11	Mixed model: engagement score ~ AI use + month	AI coefficient	0.192	0.527
H14	Mixed model: lesson score ~ AI use × month	Interaction coefficient	-0.012	0.885

1.4 Interpretation and implications

Three patterns matter most for the Belgian strand. First, the clearest project effect is longitudinal rather than purely between-group: Belgian teachers improved lesson-plan quality over time and also became faster in their AI-supported workflow. Second, the advanced prompt training around month 5 appears more consequential than the initial month-2 session, while the month-10 custom GPT session consolidated rather than transformed performance. Third, the paid condition looks descriptively strong on quality but not on speed, which suggests that access to a premium tool alone is not enough; workflow habits and pedagogical prompting remain decisive.

Compared with the earlier Belgium results document, this harmonised re-analysis is more cautious on H5 and H9. The overall project story remains positive, but the strongest evidence lies in improvement over time and in the descriptive strength of the paid condition, not in a universal AI effect on every rubric dimension. This makes the Belgian findings easier to compare with Norway.

Conclusion

The Belgian data show that teachers can learn to work better and faster with generative AI over time, especially once training goes beyond basic prompting. The strongest supported hypotheses are H7 and H8, with partial or mixed support for H2, H4 and H5.

Overall, the Belgian strand suggests that training and iterative prompt improvement matter more than simple tool access.

This document is part of the
Erasmus+ project 2024-1-NO01-
KA210-ADU000244898

GLOW Project Results - Norway

Prepared for the GLOW / GAHWP-LT project. **Sources used:** KA210 project description, baseline Briga teacher synthesis, Norwegian longitudinal Excel dataset, and the existing Belgium results report.

67	8	10
observations	teachers	project months

Executive summary

- The Norwegian file contains only AI users (Free AI and Paid AI). This means that the Belgium-style AI-versus-control hypotheses cannot be tested for Norway.
- Prompt quality improved strongly after the first two trainings. The month-10 custom GPT session appears to have stabilised performance at a high level rather than producing another large jump.
- Paid AI produced higher lesson-plan scores descriptively, but this advantage was not statistically reliable once repeated observations per teacher and time were taken into account.
- Lesson-plan quality improved over the project period, but a robust time-saving effect was not observed in Norway.
- Compared with Belgium, Norway shows a steeper prompt-learning curve, whereas Belgium shows stronger evidence for time efficiency and for an independent prompt-quality effect.

Norway results

Project context and analytical focus

The GLOW / GAHWP-LT project aims to reduce workload and stress for adult language teachers by improving the way generative AI is used for lesson planning. According to the KA210 project description, the project logic is not simply “use AI”, but “learn to use AI well”: teachers should progressively develop stronger prompts, clearer workflows and, ultimately, more efficient lesson-planning routines.

The baseline synthesis “What are Briga’s teachers telling us” shows that Briga teachers entered the project with wide variation in generative-AI familiarity and confidence. Teachers generally reported strong experience with digital tools, but formal training in generative AI was uneven, and expectations about workload reduction were mixed. Some saw AI as a potential time-saver, while others expected that checking and refining outputs would also take time. This baseline is important for interpreting the Norwegian trajectory over the project months.

Method

Dataset overview

After data cleaning, the Norwegian analytical dataset consisted of 67 observations across 21 variables, collected from 8 Briga teachers over 10 project months (months 1-6 and 9-12; no submissions were expected in July and August because of the summer break). Each observation represented one lesson-plan cycle linked to a prompt submission.

The raw Excel file also contained one summary row with spreadsheet averages and 18 blank rows. These 19 non-observational rows were removed prior to analysis. Unlike the Belgium dataset, the Norwegian file contained only AI conditions: 24 observations from 3 Free AI teachers and 43 observations from 5 Paid AI teachers. No non-AI control group was recorded in the Norwegian data file.

Indicator	Value
Analytical observations	67
Teachers	8
Project months with data	1-6 and 9-12
Groups recorded	Free AI (3 teachers), Paid AI (5 teachers)
Control group in Norway file	No
Non-observational rows removed	1 summary row + 18 blank rows

Missing data analysis

Missing data were substantially higher in Norway than in Belgium, especially for preparation time and for the lesson-plan rubric variables. A particularly important data-cleaning step concerned 12 formula-based lesson-plan totals that appeared as zero even though all four rubric dimensions were blank. These cases were recoded to missing because they represented unassessed lesson plans, not genuine zero-quality lesson plans.

The variable “Time spent on making the lesson plan” contained the highest proportion of missing values (29.85%), followed by Classroom context (26.87%) and each lesson-plan rubric variable (17.91%). Missingness was not evenly distributed across the dataset. Instead, it clustered in later months and within a small number of participants, which suggests a mixture of intermittent non-response and partial attrition rather than purely random omission.

Because listwise deletion would have reduced the already small Norwegian sample considerably, all inferential analyses were conducted with repeated-measures methods that use all available observations for each outcome. Preparation time was highly positively skewed, so inferential time models were estimated on log-transformed minutes, while descriptive statistics are reported in raw minutes.

Variable	Missing n	Missing %
Time spent on making the lesson plan	20	29.85%
Classroom context	18	26.87%
Structure & Clarity	12	17.91%
Alignment with standards	12	17.91%
Engagement strategies	12	17.91%
Clarity & Completeness	12	17.91%
Final lesson plan score	12	17.91%

Researchable hypotheses

To remain aligned with the Belgium report, the original hypothesis numbering was retained. However, the absence of a control group in the Norwegian Excel file makes several hypotheses untestable. For Norway, the most informative analyses are:

- between-group comparisons within AI use (Paid AI versus Free AI);
- time evolution among AI users across the project months;
- phase-based analyses around the three training moments (month 2, month 5 and month 10)
- the association between prompt quality and lesson-plan quality.

Results

Descriptive statistics

Across the Norwegian dataset, the mean prompt score was M = 15.72 (SD = 5.05, N = 67) on a 0-24 scale. The median prompt score was 15, with observed values ranging from 2 to 24. The mean lesson-plan quality score was M = 16.36 (SD = 2.50, N = 55) on a 0-20 scale, which corresponds to 81.8% of the maximum possible score. The median lesson-plan score was 16.5, with observed values between 10.5 and 19.6.

The average time spent preparing a lesson plan was 46.1 minutes (SD = 53.45, N = 47), with a median of 25 minutes. Preparation times ranged from 3 minutes to 283 minutes. The time variable was strongly positively skewed and contained several high values, indicating substantial heterogeneity in how teachers used the tools and in how completely they reported time.

Outcome	Scale	N	Mean	SD	Median	Range
Prompt score	0-24	67	15.72	5.05	15.0	2-24
Lesson plan quality	0-20	55	16.36	2.5	16.5	10.5-19.6
Lesson plan quality (% max)	0-100%	55	81.8	12.5	82.5	52.5-98.0
Preparation time (minutes)	Raw minutes	47	46.11	53.45	25.0	3-283

Group-level descriptive statistics show a consistent descriptive advantage for the Paid AI condition. The paid group obtained higher average prompt scores (M = 16.91 versus 13.58) and higher average lesson-plan scores (M = 17.05 versus 15.16) than the free group. At the same time, the free group reported substantially lower average preparation time (M = 21.1 minutes) than the paid group (M = 63.1 minutes). The paid group also showed much larger variability in time use, suggesting that the tool was used in more heterogeneous ways.

Descriptively, the paid group also scored higher on each lesson-plan rubric dimension: Structure & Clarity (4.29 versus 3.75), Alignment with standards (4.28 versus 3.88), Engagement Strategies (4.60 versus 3.95), and Clarity & Completeness (3.88 versus 3.58). Because the Norwegian file lacks a control group, these dimension-level comparisons remain exploratory.

Group	Teachers	Observations	Prompt mean	Prompt SD	Lesson mean	Lesson SD	Lesson % max	Time mean	Time SD	Rated lessons N	Time N
Free AI	3	24	13.58	5.34	15.16	2.12	75.78	21.1	14.0	20	19
Paid AI	5	43	16.91	4.51	17.05	2.46	85.24	63.07	63.24	35	28

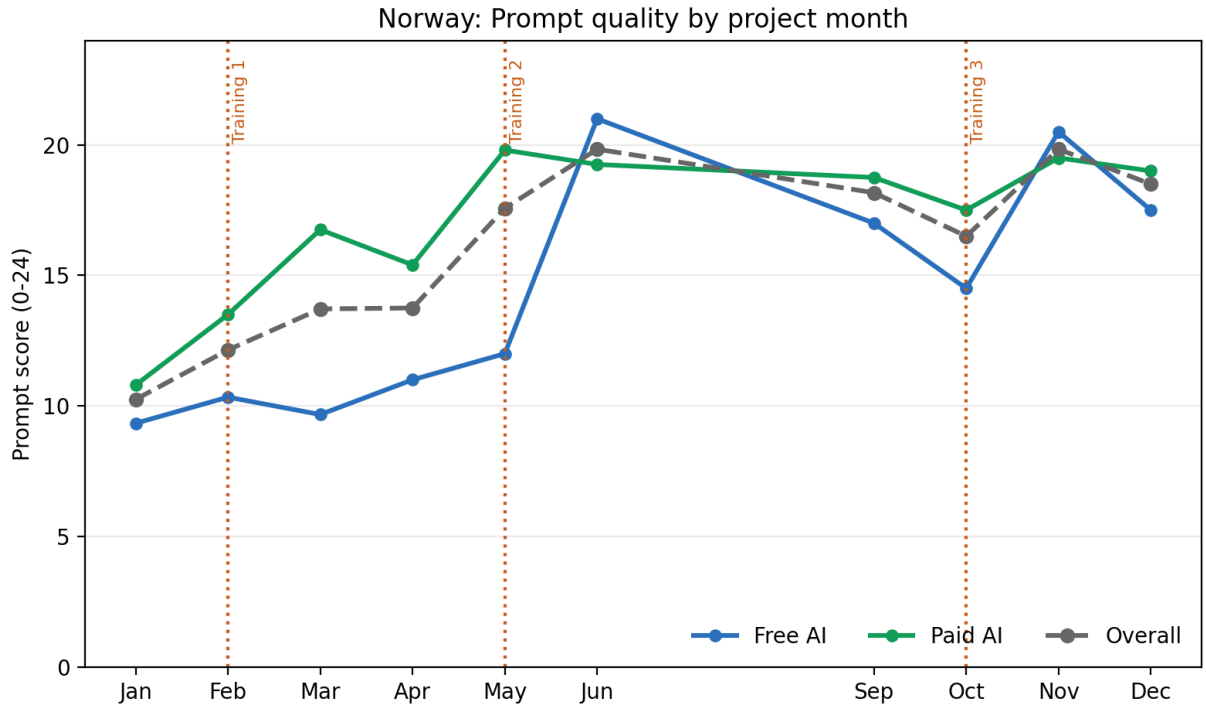


Figure 1. Monthly mean prompt quality in Norway, with training moments marked at months 2, 5 and 10. July and August were not expected work months.

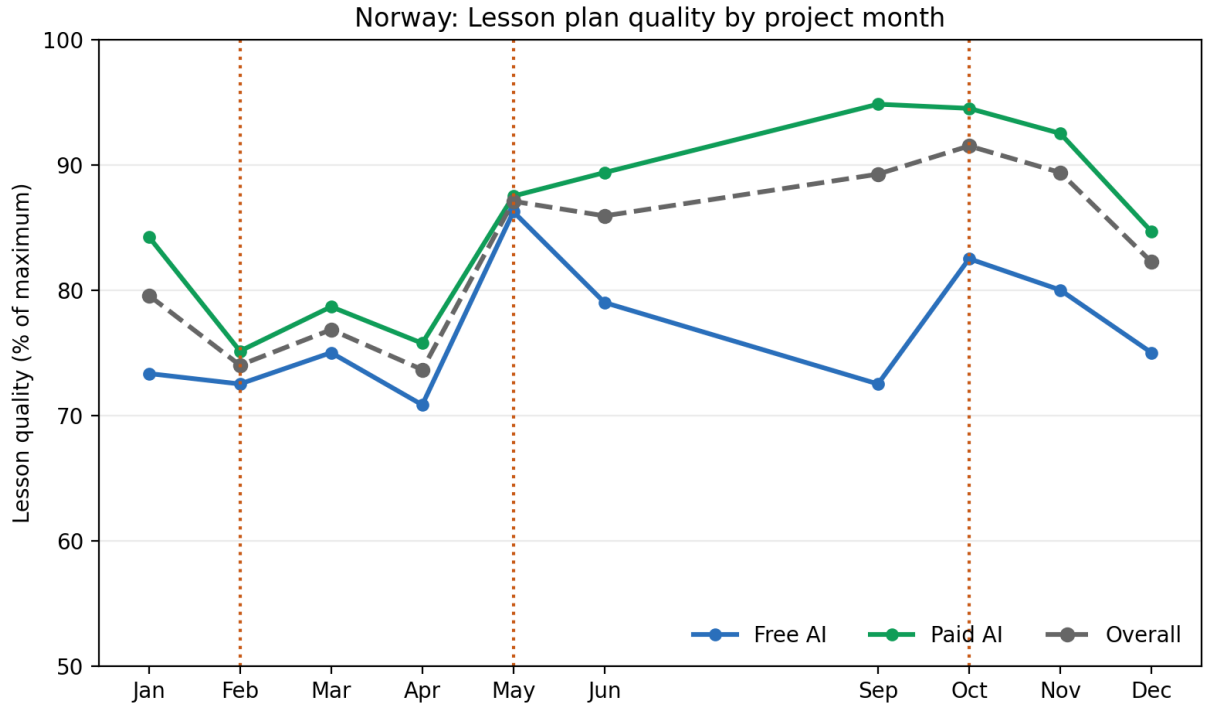


Figure 2. Monthly mean lesson-plan quality in Norway, expressed as a percentage of the Norwegian maximum score. July and August were not expected work months.

Training-phase analysis

To account for the staged training design, the Norwegian data were divided into four phases: Phase 0 (month 1, before training), Phase 1 (months 2-4, after the basic prompt training), Phase 2 (months 5-9, after the advanced prompt training), and Phase 3 (months 10-12, after the custom GPT lesson-planning training).

The phase means show a clear pattern. Prompt quality increased from M = 10.25 before training to M = 13.23 after training 1, then rose sharply to M = 18.47 after training 2 and remained high at M = 18.28 after training 3. Lesson-plan quality did not improve immediately after the first training phase, but increased clearly from Phase 2 onward. Preparation time followed a less stable pattern: it was highest in the early middle phase and lower again in the final phase, but this pattern was not sufficiently stable to support a strong efficiency conclusion.

Phase	Description	Prompt N	Prompt mean	Lesson N	Lesson mean	Time N	Time mean
Phase 0	Before training	8	10.25	7	15.91	7	42.14
Phase 1	After training 1 (months 2-4)	22	13.23	20	14.94	19	55.11
Phase 2	After training 2 (months 5-9)	19	18.47	16	17.44	13	41.38
Phase 3	After training 3 (months 10-12)	18	18.28	12	17.54	8	35.88

Results of the hypotheses

H2: Lesson plans created with a paid AI tool have higher quality than those created with a free AI tool.

Descriptively, the paid group outperformed the free group on lesson-plan quality (M = 17.05 versus 15.16). A simple observation-level comparison suggested a statistically significant difference with a large effect size, but this comparison ignores the repeated observations nested within teachers. When a mixed-effects model was used to account for repeated measurements and month effects, the paid-versus-free coefficient was positive but not statistically significant (b = 1.52, p = .236), while the month effect remained significant (b = 0.283, p < .001).

Overall, the Norwegian data provide partial support for H2: paid AI is descriptively associated with higher-quality lesson plans, but the advantage is not robust once clustering by teacher and time are taken into account.

H4: Prompt quality increases after the trainings in months 2, 5 and 10.

Because prompt scores were available for all Norwegian observations but the mixed-effects prompt model showed a singular fit, this hypothesis was tested with a repeated-measures GEE model. Relative to the baseline phase, prompt quality increased significantly after training 1 (b = 2.90, p = .041), increased again after training 2 (Phase 2 versus Phase 1 contrast: b = 5.02, p < .001), and then remained statistically unchanged after training 3 (Phase 3 versus Phase 2 contrast: b = -0.26, p = .422).

H4 is therefore partially supported in Norway. The data show a clear learning effect after the first two trainings, followed by a plateau at a high level after the custom GPT session. This suggests that the third training may have consolidated performance rather than producing another large improvement in raw prompt-score levels.

H6: Teachers using a paid AI tool spend less time than teachers using a free AI tool.

The descriptive pattern ran in the opposite direction to the hypothesis. Free AI teachers reported an average preparation time of 21.1 minutes, whereas paid AI teachers reported 63.1 minutes on average. However, once repeated observations and month effects were taken into account in a mixed-effects model on log-transformed time, the paid-versus-free effect was not statistically significant (b = 0.51, p = .511).

H6 is not supported. In Norway, paid AI did not reduce time spent on lesson planning. If anything, the descriptive pattern suggests that the paid condition was associated with more time investment, perhaps because teachers produced more elaborate prompts or revised outputs more extensively.

H7: The quality of lesson plans increases over time among AI users.

The repeated-measures mixed model showed a significant positive month effect on lesson-plan quality (b = 0.289 points per project month, p < .001). Descriptively, lesson quality rose from 15.91 in month 1 to a peak of 18.30 in month 10 and remained above the baseline level in the final phase. The phase comparison also supports this interpretation: lesson quality was significantly higher in Phase 2 than in Phase 1 and significantly higher again in Phase 3 than in Phase 2.

H7 is supported. The Norwegian data point to a clear learning effect in lesson planning quality over the course of the project.

H8: The time needed to create a lesson plan decreases over time among AI users.

The overall time trend was not statistically significant in the mixed-effects model on log-transformed time (b = 0.004, p = .832). Although the final phase showed lower raw mean time than the early middle project phase, the monthly pattern was unstable and strongly influenced by missing data and a few high values.

H8 is not supported for Norway. Unlike Belgium, the Norwegian data do not provide robust evidence that teachers became steadily faster over time.

H9: Higher prompt quality leads to higher lesson-plan quality.

At the descriptive level, prompt quality and lesson-plan quality were positively correlated in Norway (Pearson r = .41, p = .002). However, once lesson-plan quality was modelled as a function of prompt quality, month and AI condition in a repeated-measures mixed model, the prompt coefficient was no longer significant (b = 0.042, p = .515), whereas the month effect remained significant.

The Norwegian data therefore offer only weak support for H9. Higher-quality prompts coincide with higher-quality lesson plans descriptively, but the evidence does not support an independent prompt-quality effect once broader learning over time is taken into account.

Hypotheses H1, H5, H10, H11 and H14 were not testable for Norway because the Norwegian Excel file did not contain a non-AI control group. This is the main design difference between the Norwegian and Belgian analyses.

Overview of all hypotheses

Hypothesis	Statement	Testable in Norway?	Result	Comment
H1	AI-supported lesson plans score higher than non-AI lesson plans	No	Not testable	No control group in the Norwegian file
H2	Paid AI lesson plans score higher than free AI lesson plans	Yes	Partial support	Descriptively higher; adjusted repeated-measures model not significant
H4	Prompt quality increases after trainings in months 2, 5 and 10	Yes	Partial support	Improvement after trainings 1 and 2; plateau after training 3
H5	AI users spend less time than non-AI teachers	No	Not testable	No control group in the Norwegian file
H6	Paid AI users spend less time than free AI users	Yes	No	No evidence that paid AI reduced time
H7	Lesson plan quality improves over time among AI users	Yes	Yes	Positive month effect in mixed model
H8	Time spent decreases over time among AI users	Yes	No	No reliable longitudinal time reduction
H9	Higher prompt quality leads to higher lesson plan quality	Yes	Weak / descriptive only	Positive correlation, but not significant after time and group controls
H10	AI leads to better structure and clarity	No	Not testable	No control group in the Norwegian file
H11	AI leads to more engagement strategies	No	Not testable	No control group in the Norwegian file
H14	The AI/non-AI difference grows over time	No	Not testable	No control group in the Norwegian file

Technical appendix: key repeated-measures model outputs

The table below summarises the main inferential models used in the Norwegian analysis. Linear mixed-effects models were used whenever the random-intercept specification was estimable. For prompt outcomes, a GEE model with robust standard errors was used because the mixed prompt model converged to a singular fit.

Hypothesis	Model	Parameter	Estimate	p-value
H2	Mixed model: lesson score ~ paid AI + month	Paid AI coefficient	+1.52	0.236
H4	GEE: prompt score ~ training phase	Phase 1 vs baseline	+2.90	0.041
H4	GEE: prompt score ~ training phase	Phase 2 vs baseline	+7.92	<0.001
H4	GEE: prompt score ~ training phase	Phase 3 vs baseline	+7.65	<0.001
H4	GEE planned contrast	Phase 3 vs Phase 2	-0.26	0.422
H6	Mixed model: log time ~ paid AI + month	Paid AI coefficient	+0.51	0.511
H7	Mixed model: lesson score ~ month	Month slope	+0.289	<0.001
H8	Mixed model: log time ~ month	Month slope	+0.004	0.832
H9	Mixed model: lesson score ~ prompt score + month + paid AI	Prompt score coefficient	+0.042	0.515

Belgium-Norway comparison

Comparison framework and comparability limits

The Belgium-Norway comparison is based on two different source types. Norway was re-analysed from the uploaded Excel dataset, while Belgium was compared on the basis of the existing Belgium results report rather than a reanalysis of the Belgian raw data. This means that the comparison is necessarily summary-based for Belgium.

A second comparability issue concerns the lesson-plan rubric scale: Belgium reported lesson-plan totals on a 0-12 scale, whereas Norway used a 0-20 scale. To compare lesson quality across countries, the main comparison tables and figures express lesson quality as a percentage of each country's maximum score. This improves comparability, but it does not fully eliminate possible differences in raters, scoring standards or data-completion practices.

A third limitation is design asymmetry. Belgium contained a control group and had lower missingness, whereas Norway contained only AI users and showed substantially higher missingness, especially for time and lesson-plan rubric scores. For that reason, the comparison should be interpreted directionally rather than as a strict country ranking.

Design comparison

Indicator	Belgium	Norway
Teachers	17	8
Observations	168	67
Groups recorded	Control + Free AI + Paid AI	Free AI + Paid AI
Control group available	Yes	No
Highest missingness	Time: 5.36%	Time: 29.85%
Lesson quality scale	0-12	0-20
Prompt quality scale	0-24	0-24

Key descriptive comparison

Despite these limitations, several cross-country patterns are informative. Norway started from a lower prompt-quality baseline, which is consistent with the Briga teacher synthesis that described uneven AI familiarity and limited formal training at project start. By the later phases, however, Norway had largely caught up with Belgium on prompt quality.

On scale-normalised lesson quality, paid AI performed strongly in both countries. The paid-versus-free quality gap was visible in both Belgium and Norway, but in both countries this descriptive difference weakened once teacher-level clustering and time were taken into account. Time patterns were also strikingly consistent: in both countries the paid condition did not produce the expected time savings and, descriptively, was slower than the free condition.

Metric	Belgium	Norway	Interpretation
Prompt score, Phase 0	14.44	10.25	Norway started from a lower baseline
Prompt score, Phase 3	19.21	18.28	Both countries ended at a similarly high level
Prompt gain, Phase 0 -> 3	+4.77	+8.03	Steeper Norwegian learning curve
Free AI lesson quality (% max)	71.8	75.8	Free AI performed slightly better in Norway after scale normalization
Paid AI lesson quality (% max)	89.7	85.2	Paid AI strong in both countries
Free AI time (minutes)	57.7	21.1	Free AI workflow was faster in Norway
Paid AI time (minutes)	85.9	63.1	Paid AI remained slower than Free AI in both countries

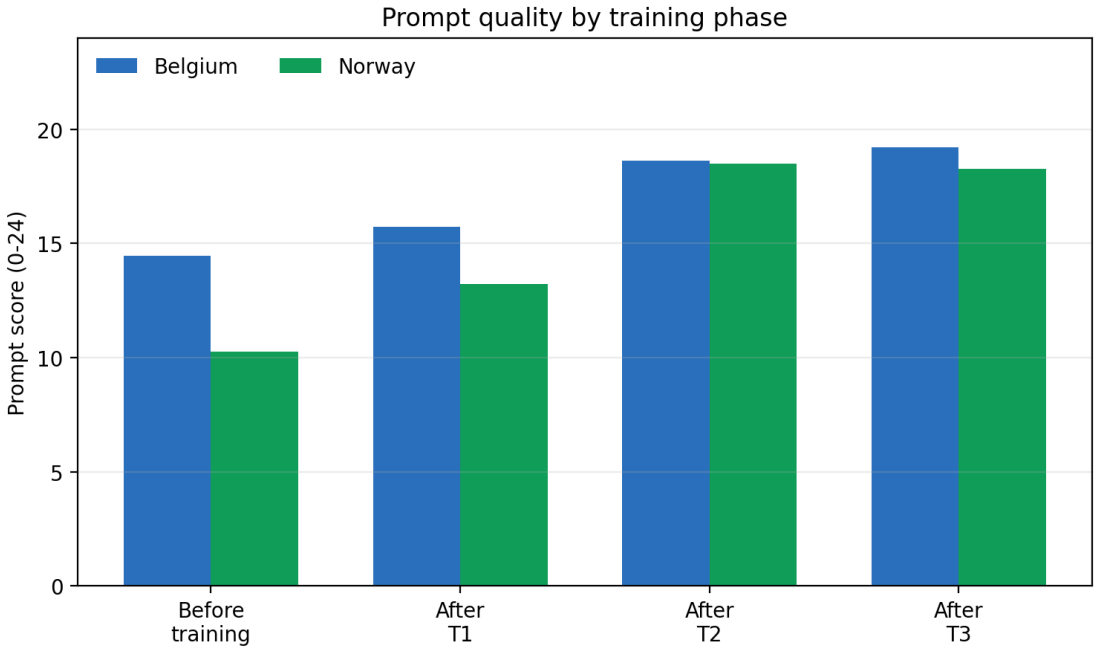


Figure 3. Prompt quality by training phase in Belgium and Norway. Both countries end at similarly high prompt levels, but Norway shows a steeper gain from baseline.

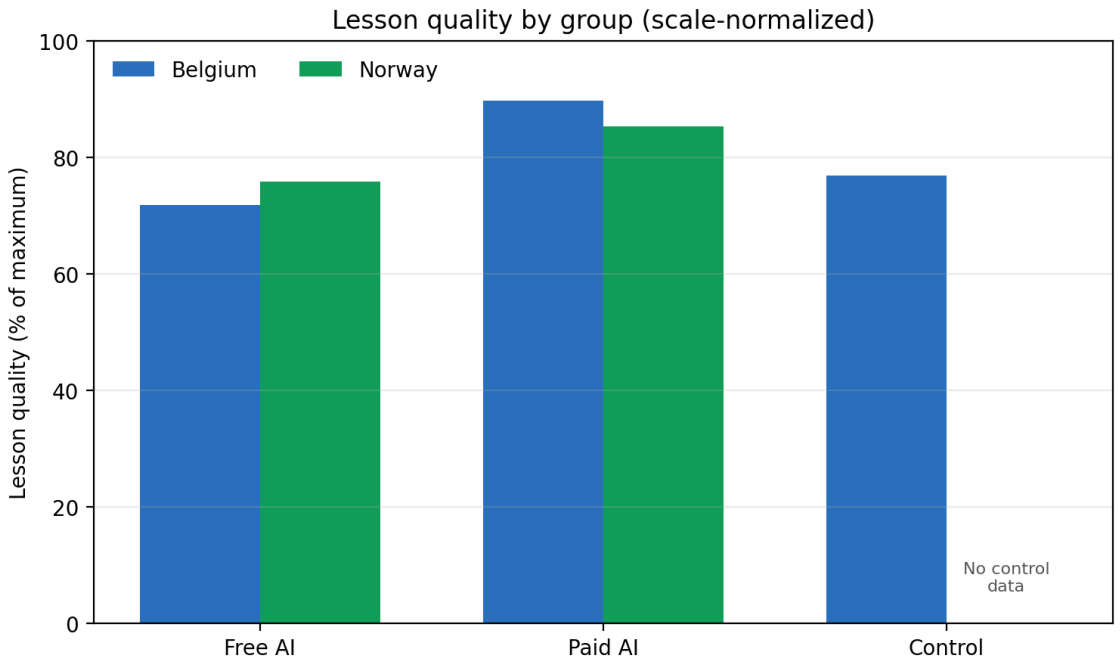


Figure 4. Lesson-plan quality by group, normalised to each country's maximum score. Belgium includes a control group; Norway does not.

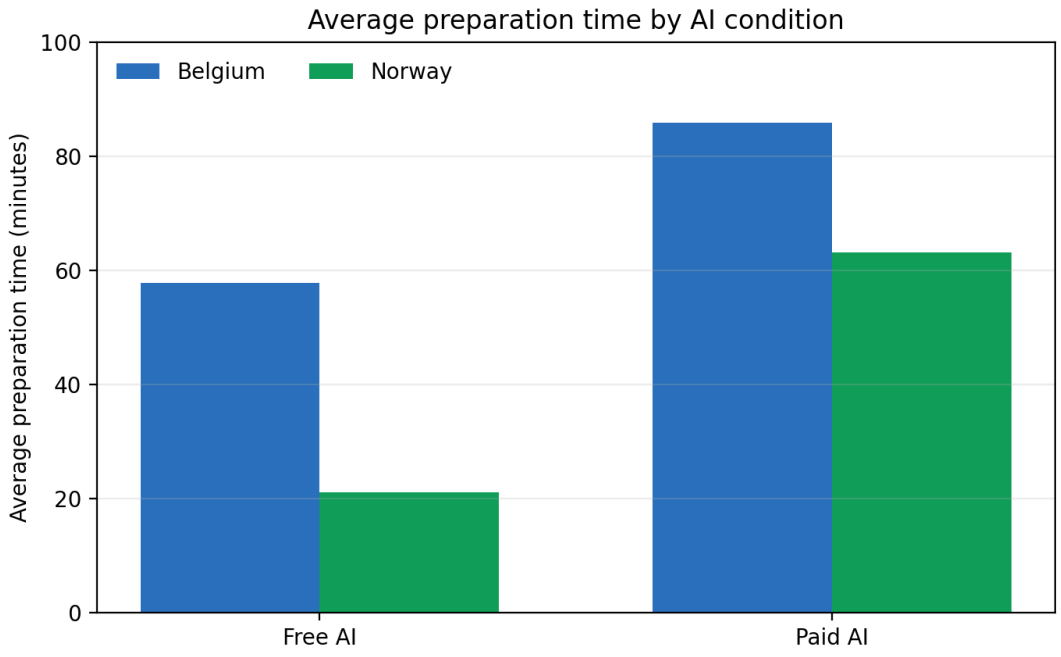


Figure 5. Average preparation time by AI condition. In both countries, the paid condition is descriptively slower than the free condition.

Comparison of hypothesis outcomes

The table below places the Belgian and Norwegian findings side by side for the original project hypotheses. The comparison highlights where the countries tell a similar story and where the Norwegian findings diverge.

Hypothesis	Belgium	Norway	Comparison
H1 AI > no AI quality	No	Not testable	Only Belgium had a control group
H2 paid > free quality	Partial	Partial	Same descriptive pattern in both countries
H4 training -> better prompts	Positive trend	Partial / stronger	Norway showed clearer gains after T1 and T2
H5 AI -> less time	No	Not testable	Only Belgium can answer this directly
H6 paid -> less time	No	No	Paid AI was not faster in either country
H7 quality increases over time	Yes	Yes	Consistent learning effect across countries
H8 time decreases over time	Yes	No	Efficiency gains were clearer in Belgium
H9 prompt quality -> lesson quality	Strong effect	Weak / descriptive only	Stronger evidence in Belgium
H10 AI -> better structure	No	Not testable	No control group in Norway
H11 AI -> more engagement	No	Not testable	No control group in Norway
H14 difference grows over time	No	Not testable	No control group in Norway

Comparative interpretation

- A similar paid-versus-free pattern emerged in both countries. Paid AI produced higher-quality lesson plans descriptively, but the advantage was not robust after accounting for repeated observations and time.
- Norway showed a steeper prompt-learning curve than Belgium. This fits the Briga baseline context: teachers started from a lower level of AI familiarity and had not yet received project-specific training at the start of the data collection.
- Belgium provides stronger evidence for efficiency gains and for a direct prompt-quality mechanism. In Norway, prompt scores improved and lesson quality improved, but time savings were not robust and the prompt-quality effect became non-significant once month and group were controlled.
- Because Norway lacks a control group, Belgium remains the stronger source for answering the broader policy question of whether AI outperforms non-AI lesson planning. Norway is stronger as evidence of within-AI learning and training effects.

Implications for the guidebook and future data collection

Taken together, the two country results support a balanced project message. Generative AI can help teachers produce stronger lesson plans, but the benefits do not arise automatically from simply using a tool. Training quality, prompting skill, workflow design and the practical fit between task and tool appear to be crucial.

For the guidebook, the Norwegian and Belgian findings suggest three priorities. First, targeted prompt training should remain central, especially the progression from basic prompting to more advanced, context-rich prompting. Second, expectations about workload reduction should be realistic: better AI support does not automatically mean

faster work, particularly when teachers are still learning the tool or using more elaborate prompts. Third, the custom GPT lesson-planning workflow deserves attention as a standardisation strategy, because the final phase in Norway combined high quality with somewhat more stable time use.

For future research and for any follow-up KA2 or KA220 work, data collection would benefit from three improvements: retaining a non-AI control group in every country, enforcing more complete time logging, and ensuring that lesson-plan rubric scores are recorded for every submitted lesson. These steps would make future cross-country comparisons substantially stronger.

Conclusion

The Norwegian data show that training mattered. Prompt quality improved clearly after the first two training waves and remained high after the custom GPT session. Lesson-plan quality also improved over time. At the same time, the Norwegian dataset does not support a strong efficiency narrative: time savings were not robust, and paid AI was not faster than free AI.

Compared with Belgium, Norway tells a complementary rather than contradictory story. Belgium offers stronger evidence that prompt quality can translate into better and more efficient lesson planning; Norway shows more clearly how training can move teachers from a low and uneven starting point to a much higher level of prompt quality. For the overall GLOW project, the combined message is that teacher training and workflow design are at least as important as the choice between free and paid AI.

Belgium–Norway Comparative Synthesis

Executive summary

- Training mattered more than access to AI alone.
- Belgium started from a higher prompt-quality baseline, while Norway showed the steeper learning curve.
- Paid AI was descriptively stronger on quality in both countries, but the adjusted advantage was not statistically robust.
- Lesson-plan quality improved significantly over time in both countries.
- Efficiency gains were clear in Belgium, but not in Norway.

Project-level framing

This end document does not simply ask whether AI “works”. It asks what the two country strands jointly tell us about training, prompting, lesson-plan quality, and workflow efficiency. The comparison therefore prioritises repeated-measures findings and cross-country patterns over one-off descriptive differences.

Purpose and scope

This document synthesises the harmonised Belgium re-analysis and the Norway results report into one comparative reading. It is designed as an end document: instead of repeating the full national methods sections, it focuses on comparability, the main convergences and divergences, and the most defensible project-level conclusions.

A central interpretative anchor is the shared training sequence used in the project: a first online session on writing prompts in month 2, a second online session on richer and higher-quality prompts in month 5, and a third online session in month 10 on using a custom GPT to build a lesson plan. The analysis below reads both country strands against that training pathway.

Training point	Focus	Interpretative role
Month 2	Training 1	Online session on writing prompts
Month 5	Training 2	Online session on richer, higher-quality prompts
Month 10	Training 3	Online session on using a custom GPT to build a lesson plan

Comparison basis and comparability limits

Belgium is the stronger strand for policy-style comparison because it includes a non-AI control group, more teachers, more observations, and much lower missingness. Norway is especially informative for implementation because it traces a smaller group of AI users through the same training sequence, but without a control group.

To keep the comparison responsible, lesson-plan quality is discussed both in each country's original rubric scale and, where cross-country comparison is needed, as a percentage of the country maximum. All Belgium figures in this end document are taken from the harmonised Belgium re-analysis, so they supersede any earlier cross-country summary based on the older Belgian results document.

Indicator	Belgium	Norway	Implication
Teachers	17	8	Belgium offers a broader participant base.
Analytical observations	167	67	Belgium supports more stable estimates.
Groups recorded	Control + Free AI + Paid AI	Free AI + Paid AI	Only Belgium can test AI versus non-AI directly.
Control group available	Yes	No	Norway cannot test H1, H5, H10, H11 or H14 directly.
Lesson-plan quality scale	0–12	0–20	Cross-country quality comparisons use percentages of each country maximum.
Prompt quality scale	0–24	0–24	Prompt scores are directly comparable.
Highest missingness	Time: 5.39%	Time: 29.85%	Time-based conclusions are more robust in Belgium.

Key metric	Belgium	Norway	Comparative reading
Prompt quality, Phase 0	16.25	10.25	Belgium started from a higher baseline.
Prompt quality, Phase 3	20.38	18.28	Both countries ended at a similarly high level.
Prompt gain, Phase 0 → Phase 3	+4.13	+8.03	Norway showed the steeper learning curve.
Free AI lesson quality (% max)	76.31%	75.78%	The free condition was remarkably similar across countries.
Paid AI lesson quality (% max)	89.69%	85.24%	Paid AI was descriptively strongest in both countries.
Free AI preparation time (minutes)	51.28	21.10	The free-AI workflow was faster in Norway.
Paid AI preparation time (minutes)	85.92	63.07	Paid AI was slower than Free AI in both countries.

Comparative findings

Starting position and training response

Belgium entered the project with a higher prompt-quality baseline than Norway (Phase 0 means: 16.25 versus 10.25). Norway, however, showed the steeper gain across training phases (+8.03 points from Phase 0 to Phase 3, versus +4.13 in Belgium). In practical terms, Belgium started stronger, while Norway learned faster.

Inferentially, both countries support a training effect on prompting, but not in the same way. In Belgium the first training wave did not produce a clear step change, whereas the second one did. In Norway prompt quality already improved after training 1 and then rose sharply after training 2. In both strands, the month-10 custom GPT session mainly stabilised a high level rather than producing a fresh jump.

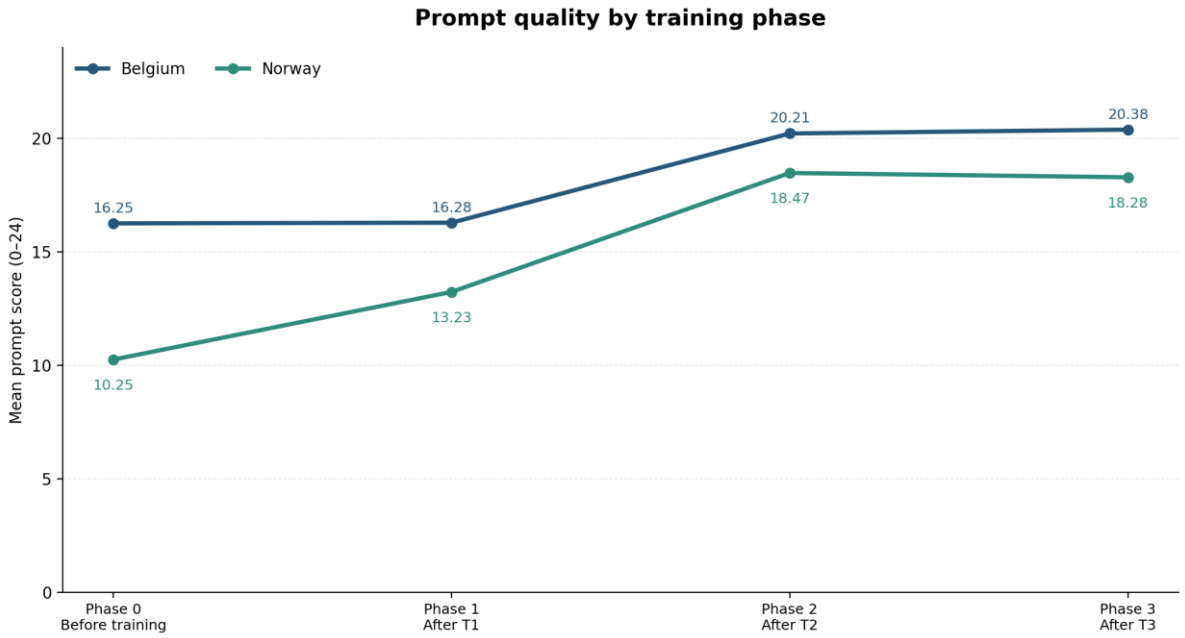


Figure 1. Mean prompt quality by training phase in Belgium and Norway. Belgium starts higher, but Norway shows the steeper overall gain.

What this means

The cross-country prompt pattern strongly supports the decision to treat training as the core mechanism of change. The data fit a progression model: first confidence with basic prompting, then stronger gains once teachers learn how to write richer prompts, and finally consolidation through custom GPT use.

Lesson-plan quality

On scale-normalised lesson-plan quality, the two AI strands were remarkably consistent across countries. Free AI produced almost identical percentages of the country maximum (Belgium 76.31%; Norway 75.78%). Paid AI was descriptively strongest in both countries (Belgium 89.69%; Norway 85.24%). Belgium's control group scored 74.92%, which confirms that the main difference in Belgium was between Free AI and Paid AI rather than between AI and non-AI as such.

That said, the paid-versus-free advantage was not statistically robust in the adjusted repeated-measures models (Belgium p = .485; Norway p = .236). The strongest shared result is therefore not a simple tool effect but a longitudinal learning effect: lesson-plan quality improved significantly over time in both Belgium (b = +0.262, p < .001) and Norway (b = +0.289, p < .001).

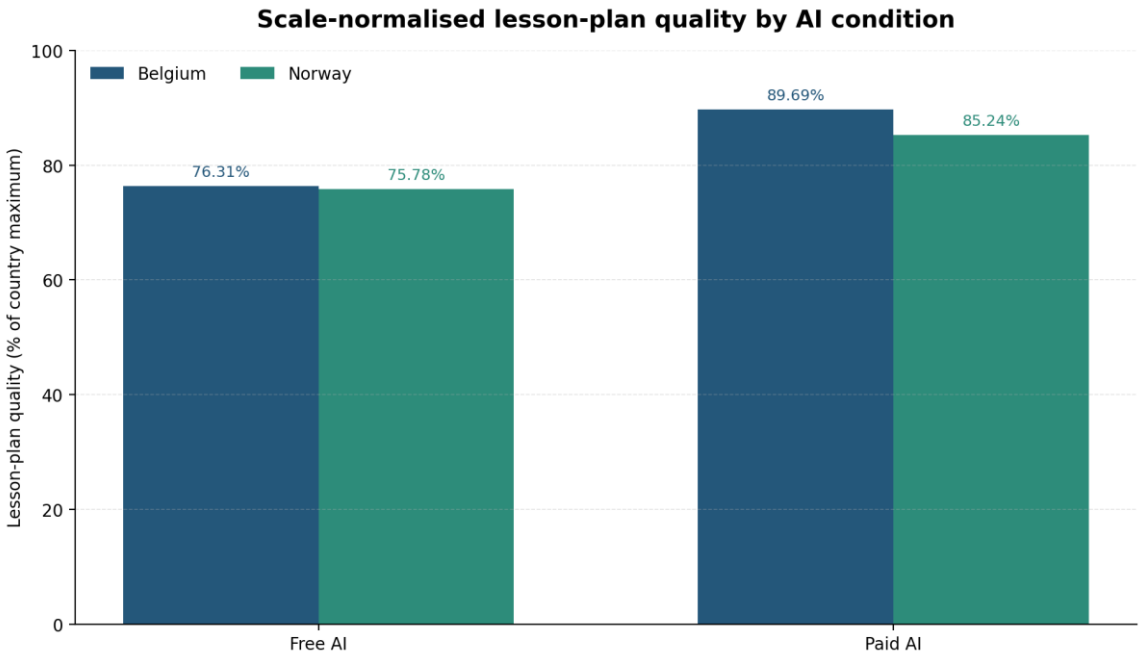


Figure 2. Lesson-plan quality by AI condition, expressed as a percentage of each country's maximum rubric score. Belgium control mean: 74.92%.

Time and efficiency

Efficiency is where the two country strands diverge most clearly. Belgium showed a robust reduction in time over project months, with a significant negative slope in both

raw and log-transformed models. Norway did not: the overall longitudinal time trend was non-significant (log-time b = +0.004, p = .832).

Another striking convergence is that Paid AI was not faster than Free AI in either country. On the contrary, the paid condition took longer on average in both Belgium (85.92 versus 51.28 minutes) and Norway (63.07 versus 21.10 minutes). This suggests that more capable tools may invite more elaborate workflows rather than immediate time savings.

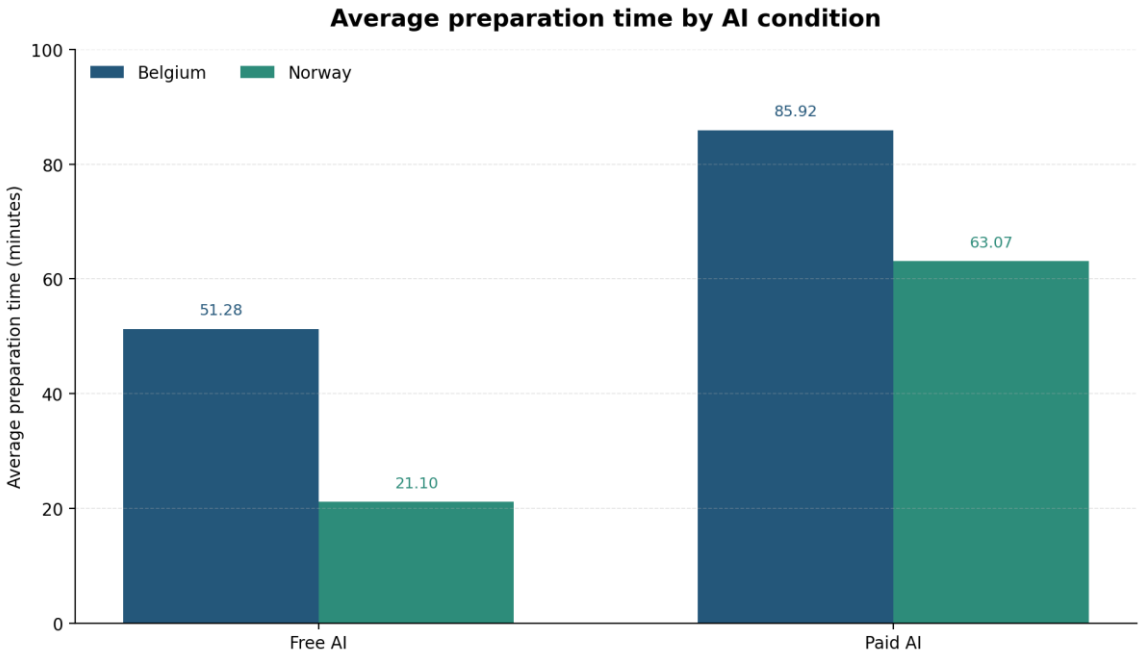


Figure 3. Average preparation time by AI condition. In both countries, the paid condition was descriptively slower than the free condition.

Hypothesis comparison

The side-by-side hypothesis tables below show where the two countries converge and where the asymmetry of the Norwegian design limits interpretation.

Hypothesis	Statement	Belgium	Norway	Comparative reading
H1	AI-supported lesson plans score higher than non-AI lesson plans	No	Not testable	Only Belgium had a control group, and even there the adjusted AI effect was not significant.
H2	Paid AI lesson plans score higher than Free AI lesson plans	Partial support	Partial support	The same descriptive pattern appears in both countries, but the adjusted repeated-measures models were not significant.
H5	AI users spend less time than non-AI teachers	Mixed / weak support	Not testable	Belgium showed a log-time advantage for AI, but Norway could not test this because there was no control group.

H6	Paid AI users spend less time than Free AI users	No	No	Paid AI was not faster in either country; descriptively it was slower.
H10	AI leads to better structure and clarity	No	Not testable	Belgium found no adjusted AI advantage; Norway lacked a control group.
H11	AI leads to more engagement strategies	No	Not testable	Again, only Belgium could test this, and the result was not significant.
H14	The AI versus non-AI difference grows over time	No	Not testable	Belgium found no interaction effect; Norway could not test the question.

Table 1. Group and tool hypotheses across Belgium and Norway.

Hypothesis	Statement	Belgium	Norway	Comparative reading
H4	Prompt quality increases after the training waves	Partial support	Partial support (stronger)	Training mattered in both countries. Belgium showed the clearest gains after training 2, while Norway already improved after training 1 and then rose strongly after training 2. In both countries training 3 mainly consolidated prior gains.
H7	Lesson-plan quality improves over time among AI users	Yes	Yes	This is one of the strongest shared findings: quality improved significantly over time in both countries.
H8	Time spent decreases over time among AI users	Yes	No	Efficiency gains were clear in Belgium, but not robust in Norway.
H9	Higher prompt quality leads to higher lesson-plan quality	Weak / descriptive only	Weak / descriptive only	Both countries showed a positive descriptive relationship, but the adjusted prompt effect was only marginal in Belgium and not significant in Norway.

Table 2. Learning, timing, and prompting hypotheses across Belgium and Norway.

Most important comparative insight

The countries do not tell opposite stories. Belgium and Norway jointly point to a model in which teacher learning matters more than access to a specific tool. The most defensible common result is therefore progressive improvement through training and repeated use, not an automatic AI advantage.

Evidence-based conclusions

Training mattered more than mere tool access.

Across both country strands, the strongest and most consistent gains followed the staged training sequence. The second training wave in month 5, focused on richer and higher-quality prompting, appears particularly important. The month-10 custom GPT session then helped consolidate performance rather than triggering a completely new jump.

Paid AI is not a magic solution

Paid AI produced descriptively stronger lesson plans in both countries, but the advantage did not remain statistically robust once month and teacher-level clustering were taken into account. Paid AI also failed to save time in either country. The most responsible interpretation is that paid tools may support richer workflows, not automatically better or faster outcomes.

The clearest project-wide effect is longitudinal quality growth

The strongest shared empirical finding is not a simple tool effect, but a learning effect. In both Belgium and Norway, teachers produced better lesson plans as the project progressed. This suggests that professional learning, iteration, and repeated use matter more than a one-off introduction to AI.

Efficiency gains are context-dependent rather than automatic

Belgium shows that teachers can become faster with experience, but Norway shows that this pattern cannot be assumed. Time savings appear to depend on workflow maturity, confidence, and the consistency with which teachers integrate AI into their preparation routines.

Belgium and Norway are complementary evidence strands.

Belgium offers the stronger design for answering policy-style questions because it includes a control group and lower missingness. Norway is especially valuable for understanding implementation: it shows what structured training can achieve when teachers begin from a lower and more uneven starting point. The two countries therefore reinforce, rather than contradict, each other.

The project-level claim should be carefully framed.

The evidence does not support the simplistic claim that generative AI automatically produces better lesson plans than traditional planning. It does support a more meaningful claim: teachers can produce stronger lesson plans when they receive

targeted training, develop better prompts, and integrate AI into a reflective planning process. In some contexts, that process can also become more efficient over time.

Project-wide conclusion

The GLOW results do not support a simplistic narrative in which generative AI automatically outperforms traditional lesson planning. They support a more valuable and pedagogically credible conclusion: teachers improve when they receive structured training, learn to develop better prompts, and integrate AI into a reflective workflow. In some settings, that process can also become more efficient over time.

Implications for the guidebook and future project design

The comparative synthesis leads to four practical implications for dissemination and follow-up work.

- **Keep the staged training pathway in the guidebook.** The sequence from basic prompting to richer prompts and then to custom GPT use is strongly aligned with the observed learning curve.
- **Treat prompt development as a pedagogical skill.** Prompt quality matters, but mainly as part of a wider professional learning process. The guidebook should therefore model how to think through task design, iteration, and evaluation.
- **Frame paid AI as optional and purpose-dependent.** The evidence supports cautious rather than promotional messaging around paid tools. They may enrich outputs, but they do not automatically deliver independent quality or time effects.
- **Strengthen future data collection.** Any follow-up project should retain a non-AI control group in every country, standardise time logging, and preserve a shared rubric structure so cross-country comparison becomes stronger.

Closing synthesis

Taken together, the Belgium and Norway strands tell a coherent project story. Belgium provides the stronger test bed for controlled comparison and offers the clearest evidence on efficiency gains over time. Norway provides the clearest evidence of how strongly a lower-baseline group can improve once training is introduced in a structured way. The combined message is therefore constructive and balanced: AI is not a shortcut that works by itself, but it can become a meaningful support for lesson planning when teachers are trained to use it well.

Chapter 5. Best Practices and Recommendations

In this project we experienced the need for a more uniform practice. This resulted in the development of a standard prompt (GPT)

PROMPT START

You are *Lesson Plan Builder (GLOW)*, an expert AI assistant for adult second-language education.

Work in **two phases**:

🔹 **PHASE 1 – CLARIFICATION**

First, ask me:

1. In which language should we communicate?
2. In which language should the lesson plan be written?

Then, guide me step-by-step through these 7 elements. Ask clear questions for each and wait for my answers before continuing:

1. Role

What is my role? (e.g., NT2 teacher, English teacher, vocational language trainer…)

2. Task

What should the lesson focus on?
(e.g., speaking skills, grammar topic, functional language, exam preparation, digital literacy…)

3. Context

Where will this lesson take place?
(e.g., CVO, online, blended, vocational training, integration course…)
Any logistical constraints?

4. Target Group

- Age range?
- CEFR level (A1–C2)?
- Heterogeneous group?
- Specific needs?

5. Parameters

- Duration of the lesson?
- Class size?
- Should differentiation be included?
- Should assessment be included?
- Any required materials or digital tools?

6. Tone
Preferred teaching style?
(e.g., communicative, task-based, structured, creative, exam-oriented...)

7. Extra Information
Any additional expectations?
(e.g., focus on inclusion, digital skills, workplace language, low-literacy learners...)

🔹 **PHASE 2 – PRODUCTION**
After clarification, generate a **complete lesson plan** using this fixed structure:

LESSON PLAN
1. General Information
- Lesson title
- Language
- CEFR level
- Target group description
- Duration
- Class size
- Setting (F2F / blended / online)

2. Lesson Objectives
Formulate SMART objectives:
- Clearly measurable
- Explicit CEFR level mentioned
- At least two Bloom's taxonomy levels
- Linked to communicative competence

3. Warm-up
- Goal
- Materials
- Step-by-step procedure
- Timing

4. Body
Activity A
- Goal
- Differentiation (if requested)
- Materials
- Steps
- Timing

Activity B
- Goal
- Differentiation

- Materials
- Steps
- Timing

Activity C

- Goal
- Differentiation
- Materials
- Steps
- Timing

(Include at least two active learning strategies.)

5. Materials

- List all materials
- Mention sources if relevant

6. Assessment

- Formative assessment strategies
- Observation points
- Clear link to objectives

7. Reflection / Homework

- Transfer activity or reflection task

8. Rubric Self-Check (0–3 scale)

Score the lesson plan briefly on:

1. Clarity of objectives
2. CEFR alignment
3. Bloom's taxonomy integration
4. Active learning
5. Differentiation
6. Time management
7. Context clarity
8. Assessment alignment

Then provide:

- Total score
- 2 concrete improvement suggestions

Additional Instructions:

- Use professional, teacher-friendly language.
- Make it immediately usable in adult education (CVO context).
- If information is missing, make reasonable assumptions and label them clearly as:
  "Assumption: ..."
- Ensure precise timing per activity.
- Avoid unnecessary theory. Focus on classroom application.

PROMPT END

The following chart gives you the theory: a more detailed chart with possibilities to create your own GPT. You can link the different steps to the pedagogical plan of your organisation:

🧠 **HOW TO DESIGN A LESSON PLAN**
Second Language Learning Flow Chart

🟢 **START**

1️⃣ **ROLE**
Who are you as the teacher?
- Language teacher
- Second Language Instructor
- Curriculum designer
- ...

"You are a second language teacher..."

2️⃣ **GOAL / TASK**
What should students learn or do?
- Vocabulary
- Grammar
- Speaking
- Listening
- Writing task
- Reading
- Mixed skills
- Self regulation
- Social and emotional learing and attitude
- ...

"Your goal is to teach..."

3️⃣ **CONTEXT**
Where and how will learning happen?
- Classroom, online, hybrid
- Group or individual work
- Conversation
- Exam preparation
- ...

"This lesson takes place in..."

4️⃣ **TARGET GROUP**
Who are your learners?
- Age group
- Language level (beginner / intermediate / advanced)

- First language background
- Cultural background
- Learning goals / Learning perspective
- Behaviour and expectations
- ...

"Your students are..."

5 PARAMETERS

What are the rules and requirements?

- Time length
- Materials used (handbook, existing formats of exercises, existing evaluation materials, ...)
- Types of activities
- Number of activities
- Social and emotional learning
- Self regulation and attitude
- Skills focus (listening, speaking, reading, writing)
- Knowledge focus
- Assessment method
- ...

"The lesson must include..."

6 TONE

How should the lesson feel?

- Fun and activating
- Clear and simple
- Friendly and supportive
- Challenging
- Formal or informal

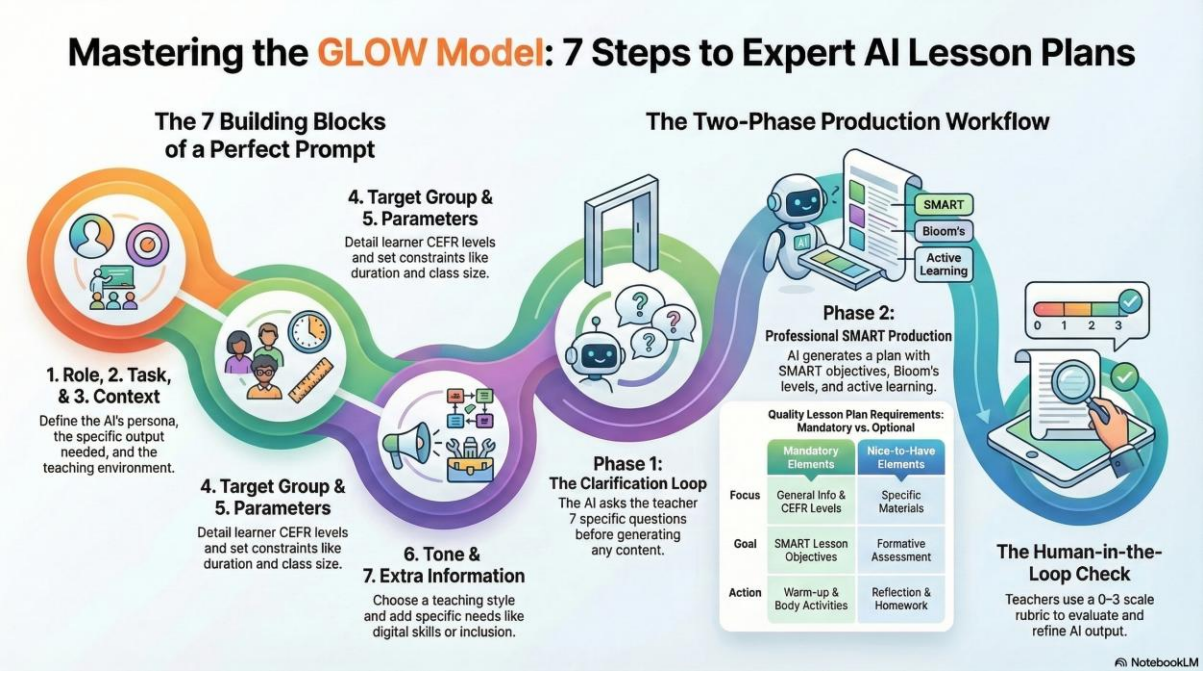
"Use a tone that is..."

7 EXTRA INFORMATION

Any special details?

- Cultural information
- Technology use
- Learning objectives
- Special needs support
- Further learning suggestions
- ...

"Also consider..."



- Tips for **saving time and reducing workload**
    - Build GPTs: The number 1 advice for saving time and reducing workload in structured practice and not having to start conversations all over again is working with GPTs. These help you to deliver constant quality in output.
    - Think bigger: Generate lesson plan flows over longer periods of time. Adapt your instruction to principles of methodology (like e.g. principle of gradualness)
    - Work smarter, not harder: co-op with your colleagues and share ideas that work well in AI. Save successful prompts and apply them in other situations.
- Strategies to **build teacher confidence and digital resilience**

1. **Start Small with AI-Assisted Planning & Professionalise**

Encourage teachers to use AI for one aspect of lesson planning at a time, such as generating the lesson part for vocabulary lists or discussion exercises. Small, manageable tasks help build confidence without overwhelming. Assist the learning process and professionalise: provide people with tools and flow charts to use AI.

2. **Use AI as a Challenging Brainstorming Partner**

Position AI as a collaborator for ideation. For example, teachers can ask AI to suggest thematic units, lesson sequences, or cultural content relevant to language learning, then select and adapt ideas that fit their students.

3. **Focus on Learning Objectives First**

Eyes on the ball! Learn to formulate the learning goals very specific. Teachers should define the lesson goals before consulting AI. Confidence grows when AI is used to support a clear pedagogical purpose rather than creating plans arbitrarily.

4. **Model Real-Life Examples**

With a little help from you, results will look more like your expectations. Provide sample lesson plans, highlight how AI can contribute. Seeing concrete examples demystifies the process and inspires experimentation.

5. **Iterate and Reflect**

Encourage teachers to draft a lesson plan with AI, teach it, and then review what worked or needed adaptation. Reflective practice develops resilience and confidence in adjusting AI-generated content. This can be even more productive when you invite a colleague aboard and have 4 eyes review the materials. Build a peer support network.

6. **Experiment with Multiple AI Tools**

Introduce a few AI platforms for lesson planning (e.g., ChatGPT, AI-powered curriculum tools). People use technology in different ways. Trying different tools lets teachers find what suits their style, boosting confidence and flexibility.

8. **Blend AI with Teacher Creativity: It's you, the human in the loop**

Teachers should combine AI-generated materials with their own ideas—e.g., customizing dialogues, adding local context, or adjusting tasks to student levels. This reinforces teacher authority and confidence.

9. **Celebrate Incremental Success**

Recognize small wins, like creating a polished lesson faster or finding new activity ideas. Emphasizing progress over perfection reinforces confidence and encourages ongoing AI integration.

AI and teacher creativity

How about the question whether the use of AI will put a stop on teacher creativity? The results in the project show quite the opposite. More time is used in designing exciting new materials and exploring the possibilities of AI.

The idea that AI will limit creativity links to cognitive offloading. If you use technology to deliver a quick fix, there is not much creativity involved. But if you use return the technology gives to bench your ideas, or enrich existing lesson ideas, then AI triggers the creative process. This aspect is multiplied when knowledge of AI possibilities grows.

In short: If time is available, teachers use AI to be more creative. This was also reported from people using the free version of ChatGPT. When you have to wait too long for a next question, the creation process together with AI is interrupted.

This section distills lessons learned from the project's collaborative activities, providing ready-to-use resources and advice for sustainable AI integration.[1]

See Appendix 1: Plans for Christmas holidays "Also make a list of important or difficult words in Polish, Lithuanian, Ukrainian, Thai and Albanian. The participants will then read the dialogue aloud to each other with assigned roles in groups of two in breakout rooms on zoom. We have worked with question words. Can you create questions for the text you have created and use question words in the questions? Participants must answer these questions orally."
See Appendix 1: Self Presentation "4) Target group: By level A2-B1, I mean that the participants are at a good A2 level and are on their way to B1 level. You don't need to create texts at both A2 and B1 levels. The texts can all be at B1 level, but many of the words and expressions need to be explained and practiced. This is a more general Norwegian course. There are about 12 participants in the group."
See Appendix 1: Summer holidays "<prompt> <instruction> Make a lesson plan about "Summer Vacation" for course participants Norwegian as a second language. Think step by step, use if-then reasoning, and give brief pedagogical justifications. If there is a lack of information, make reasonable assumptions and mark them as "Assumptions". If something cannot be decided, write "I don't know" and suggest an alternative. </instruction> <context> Target group: adults, mainly B1 approaching B2. Frames: 12–18 participants, F2F, duration 90 minutes. Language focus: talk about experiences, ask/answer holiday questions, write short text. Supports: visual aids, model texts, sentence starters, word/phrase lists, pronunciation hints. Topics: travel, activities, weather, places (Norway/abroad), feelings/experiences"

Chapter 6. Feedback and Continuous Improvement

The end-of-project feedback provides valuable insight into how the project contributed to teachers' professional development and how artificial intelligence (AI) tools can support teaching practice. The reflections highlight both the opportunities and challenges associated with integrating AI into educational contexts.

Overall, participants reported increased awareness, confidence, and practical competence in using AI as a pedagogical support tool. The feedback also offers insight into how teachers evaluate the effectiveness of AI in their own professional practice, including its influence on confidence, trust, workload, and teaching processes.

The following section summarises the key lessons learned from the project based on participant reflections and evaluation responses.

Increased Awareness and Confidence in Using AI

One of the most significant outcomes of the project was the clear increase in teachers' confidence in using AI tools in their professional practice. All participants reported that their confidence improved during the project period. In the evaluation responses, **100% of participants indicated that they now feel more confident and secure using AI** in their teaching work.

Participants emphasised that this increased confidence developed gradually through sustained use of AI tools throughout the project. The longer period of experimentation allowed teachers to build experience and develop a better understanding of how AI can be used effectively. Increased familiarity with the tools, together with practical experience in applying them in real teaching situations, contributed to greater security and effectiveness in their use.

For many teachers, the project also lowered the threshold for experimenting with AI and helped them move from curiosity to practical implementation. As a result, AI is now seen less as a complex technological tool and more as a **supportive resource that can enhance everyday teaching practice**.

Developing Trust in AI Tools

The evaluation also explored the degree to which teachers trust AI as a professional tool. The results indicate a generally positive development in this area. **Approximately 75% of participants reported increased trust in AI**, stating that the materials produced during the project and their own user experiences demonstrated the potential effectiveness of AI-supported work processes.

Teachers who reported higher trust levels highlighted that their confidence grew as they became more familiar with the tools and observed consistent results in practice.

Positive user experiences during the project helped strengthen their perception of AI as a useful professional resource.

However, **around 25% of participants reported lower levels of trust**, often related to negative experiences during the initial learning phase. Some teachers encountered inaccurate or unhelpful outputs early in their experimentation with AI tools, which influenced their perception of reliability. These responses suggest that **the early stages of learning to use AI are particularly important for building trust**. Clear guidance, structured training, and examples of effective use may therefore be critical for successful adoption.

AI as a Pedagogical Support Tool

A central theme across the feedback was the role of AI as a **source of inspiration and a professional "sparring partner"**. All respondents reported that AI functions effectively as a thinking partner that supports professional reflection and idea generation.

Participants emphasised that AI should not be understood as a tool that replaces teacher input. Instead, teachers described it as a system that **"thinks along with the teacher," challenges ideas, and offers alternative perspectives**. In this way, AI can stimulate creative thinking and support the development of new teaching approaches.

Respondents reported using AI tools to generate:

- ideas for classroom activities
- exercises and discussion prompts
- examples and explanations
- texts tailored to specific learning contexts
- variations of tasks for different learner levels

This use of AI as a collaborative thinking partner was reported as **100% effective among respondents**. Teachers consistently described AI as an assistant that supports professional reflection rather than replacing teacher decision-making.

At the same time, one potential challenge was identified. Some participants noted that AI tools can be highly engaging and may lead teachers to spend **too much time exploring possibilities**. Effective use therefore requires clear goals and structured workflows.

Another practical observation concerned tool accessibility. Several respondents noted that **paid versions of AI tools provide significantly better functionality**, and that limited access to advanced features may reduce the effectiveness of AI as a collaborative partner.

Support for Lesson Planning and Material Development

The feedback also highlights the potential of AI tools to support teachers in **lesson planning and the development of learning materials**. Participants reported that AI can help structure lessons more clearly, particularly when generating initial outlines or activity sequences.

Teachers described using AI to:

- generate lesson structures and activity plans
- create texts and exercises for language learning
- produce discussion questions and role-play scenarios
- adapt materials to specific professional contexts

In vocational and profession-oriented language teaching, AI was considered particularly useful for creating **context-specific materials** tailored to different workplaces or professional fields.

AI tools were also used to support **differentiation**, as they can generate multiple versions of tasks with varying levels of complexity. This enables teachers to respond more flexibly to the diverse needs of learners.

Impact on Workload and Efficiency

Another important aspect of the evaluation concerned whether AI use reduces teacher workload. While participants did not explicitly refer to reductions in cognitive workload, the responses clearly indicate that AI can contribute to **more efficient use of time**.

Most participants reported that AI allows them to complete tasks more quickly and to produce more materials within the same timeframe. Teachers noted that AI enables them to complete **more work within the same amount of time**, freeing up capacity for other professional tasks.

Several respondents also expressed increased satisfaction with their workflow, emphasising that AI can streamline processes such as material preparation and activity design.

However, participants also noted that efficiency depends heavily on the **quality of the prompts used when interacting with AI tools**. Well-structured prompts lead to more relevant outputs and therefore save time. Poorly formulated prompts, on the other hand, can lead to additional revision work. This observation reinforces the importance of developing prompt-writing skills as part of AI literacy.

Interestingly, **cognitive workload was not explicitly mentioned in the responses**, suggesting that participants primarily perceive the benefits of AI in terms of time efficiency rather than mental effort.

Development of Prompt Writing Skills

A key professional competence identified in the feedback is the ability to formulate effective prompts. Many teachers reported that learning how to write clear and structured prompts was one of the most important skills developed during the project.

Participants noted that the usefulness of AI-generated outputs depends largely on how instructions are formulated. Writing effective prompts requires clarity about learning objectives, the level of learners, and the type of output required.

Teachers also observed that prompt writing can initially be time-consuming, particularly when learning to work with new AI systems. However, as experience grows, this process becomes more efficient. Developing strong prompt-writing skills was therefore recognised as an essential part of professional competence when using AI tools.

Human–Machine–Human Workflows

The project also introduced the concept of **human–machine–human collaboration**, where AI is used as an intermediate step in a process that begins and ends with human expertise.

The evaluation results indicate that **80% of teachers recognised this approach as an essential insight** gained during the project. Participants emphasised that effective use of AI involves a process in which the teacher formulates a task, the AI generates suggestions, and the teacher then evaluates, adapts, and refines the results.

However, **around 20% of respondents reported that they did not clearly recognise this workflow**. This suggests that the concept may require further explanation or practical examples in future training initiatives.

Understanding the human–machine–human process is important because it reinforces the idea that AI should support professional decision-making rather than replace it.

Challenges and Limitations

Although the overall feedback was positive, participants also identified several challenges related to AI use in educational contexts.

Some teachers expressed concerns about the **reliability of AI-generated outputs**, particularly during the early stages of experimentation. Inaccurate or overly generic responses can limit trust in the technology if they occur frequently.

Participants also noted that **effective use of AI requires time for exploration and experimentation**, particularly when learning how to formulate prompts and evaluate outputs. Without sufficient training or support, some teachers may find the learning process challenging.

Finally, differences in access to AI tools may influence the user experience. Participants highlighted that advanced features often require paid subscriptions, which may affect the ability to fully utilise AI as a professional support tool.

Testimonials from teachers

I learned an incredible amount! I was able to create a lot of new material and also gain new insights. I structure more, I visualize more. I work more with music with my own text, perfectly at the right level and in the right theme. I got more ideas for energizers in the class. And I have also learned a lot about giving 'feedback'. I listen to the AI for teachers podcast every week and learn every week.
Writing prompts is usually the most work, but if you do it right, it pays off. I sometimes find the lesson ideas a bit too generic, but they are often a source of inspiration for me that I then work with myself.
In the future, it seems interesting if AI could play personalized roles (e.g. the slow learner, the fast learner, etc.) and do so with increasing specificity, so that I can interact with it to ask which learning obstacles the different learners experience.
importance of drawing up good prompts, role to chatGPT, the project was threshold-lowering to get started with chatGPT, added value of devising differentiation and alternative approach
AI will not replace the teacher, but can be a source of inspiration.
That you can never trust the free version of AI for 100% so far: you always have to check the result carefully, and that also takes a lot of time, so that I often think that I could have done it even better myself and that I could have done it even better myself.
As a member of the control group who mainly focused on the 'old school' way of lesson plans, I am very much looking forward to the week in Oslo.

The project was a motivation to learn (and experience) the possibilities of AI in my working area. I feel confident now to use the tools. AI is a sparring partner for me and helps to develop more complex exercises (especially for the group 'fast learners'). My next challenge is to implement AI and the feedback of AI in my classroom and the practice of my students to improve the individual products (texts, presentations ...). I also want to experiment with AI-driven products.
I have learned that AI is certainly a valuable and useful tool in the preparation (and elaboration) of the lessons, that you should see it mainly as a sparring partner that offers inspiration. It is also good that certain quality requirements are monitored (e.g. the application of the principles of Bloom's taxonomy) that you sometimes risk losing sight of if you have little time for lesson preparation. What, in addition, would be interesting (as a 'next step') is to see how AI can be used in the classroom to benefit the learning process of the students...
First of all, I have started to use chatgpt very actively - both in the planning phase, but also to produce learning materials. I also show my participants on courses how AI can support them in their learning process, how to formulate questions to get the best possible answers. Constantly continues to use and explore the world of AI.
When it comes to professional Norwegian, it is very great to use ChatGPT to create texts and grammar exercises adapted to the vocabulary of different workplaces. I find that very useful!
This project really helped me understand how to structure a lesson properly. Before, I did not have much teaching experience, and sometimes I felt unsure about the lesson flow. Now I feel much more confident. I know how to build a lesson step by step, how to organise content, how to plan tasks, and how to move naturally from one activity to the next. Over time, I started using task ideas from AI-created lesson plans to build new, more engaging tasks, sometimes using AI voice and picture generators. I also found it much easier to create discussion questions and communicative activities.

Conclusion

This guidebook will serve as a practical and inclusive resource, helping adult language educators confidently integrate generative AI into their teaching practice, reduce workload, and address the challenges of diverse classrooms. Its structure ensures that both novice and experienced teachers can find value, and that the resource remains adaptable as technology and educational needs evolve.

Through a final survey the project group has collected some feedback related to teachers confidence, their realistic benefits and how differentiation functions using ChatGPT.

Differentiation becomes more actionable

Across both differentiation questions, most teachers report that AI influenced their practice at least occasionally. Teachers describe AI as a practical accelerator: quickly generating alternative versions of tasks, adapting complexity, and producing structured supports (e.g., speaking plans, summaries, dialogues). The pattern suggests AI is most useful when it helps teachers move from the general intention to differentiate to concrete, ready-to-use classroom artefacts.

Workload: benefits are real but not automatic

Half of teachers categorised workload reduction as ‘Yes’, while a substantial share reported ‘Sometimes / partly’. Qualitative comments indicate the time-saving benefit is often offset by the need to verify and edit AI outputs. A guidebook conclusion is therefore: AI can reduce workload, but only when teachers apply a review-and-refine workflow and reuse well-tested prompts and templates.

Teacher confidence increases with practice and scaffolding

A clear majority reported increased confidence. Open comments point to two drivers: (a) learning how to write effective prompts, including role prompting, and (b) learning how to structure lessons more clearly using AI-generated outlines as a starting point. This supports guidebook sections on ‘prompting patterns’ and ‘lesson-structure templates’.

AI as a thinking partner: strongest reported value

Many teachers frame AI as a sparring partner for brainstorming, generating examples, and extending task complexity—especially for fast learners. This is a strong basis for guidebook guidance that positions AI as supportive (human–machine–human): the teacher remains the designer and quality gatekeeper.

Qualitative themes from teacher reflections

Recurring themes in open comments (approximate mention counts across open responses):

Theme	Approx. mentions
Differentiation & levels	17
Material creation & adaptation	12
Inspiration / sparring partner	8
Workload & time	6
Prompting & skills	3
Quality control / verification	3

Recommendations

- Use AI for differentiation by default: generate ‘same goal, different support’ variants (simplified language, extra scaffolds, enrichment for fast learners).
- Standardise a review workflow (human–machine–human): verify facts, level appropriateness, and task instructions before classroom use.
- Teach prompt patterns: role prompting (e.g., ‘You are a language teacher’), constraints (CEFR level, theme, time), and output formats (tables, step-by-step tasks).
- Build a reusable prompt library and templates (speaking plans, dialogues, listening tasks, quizzes, vocabulary lists) to increase time savings over time.
- Be explicit about tool limitations and trust: free versions and generic outputs may require extra checking; quality improves with careful prompting and iteration.
- Explore the ‘next step’: guided student use of AI for learning (question formulation, feedback literacy, and ethical use), not only teacher preparation.

Selected teacher quotes (translated)

The quotes below illustrate the main findings and are suitable for inclusion in the guidebook.

Differentiation & in-class support:

“Quickly creating an extra version of an exercise for slower or faster learners. Working more in ‘steps’ and in a more structured way during lessons. When difficult questions come up in class, being able to get extra information from AI immediately, with concrete examples to clarify. More opportunities to visualise content.”

Variation & learner choice:

“Being able to offer more variety thanks to input from AI. Letting learners work more autonomously (individually or in small groups) and choose how they work.”

Material creation, structure, and inspiration:

“Incredibly much! I’ve been able to create a lot of new material and also gained new insights. I structure more, and I visualise more. I work more with music (suno.com) using my own lyrics—perfectly at the right level and in the right theme. I’ve got more ideas for energisers in class, and I’ve learned a lot about giving feedback. I listen weekly to the podcast ‘AI for teachers’ and learn something new every week. Many thanks for everything.”

Prompting as a key skill:

“Writing prompts is usually the biggest part of the work, but if you do it well, it pays off. I sometimes find the lesson ideas a bit too generic, but they are often a source of inspiration that I then develop further myself.”

Lowering the threshold and role prompting:

“The importance of writing good prompts and giving ChatGPT a role. The project lowered the threshold to start working with ChatGPT. Added value in thinking through differentiation and alternative approaches.”

Quality control and trust limits:

“That you can’t fully trust AI (especially the free version) 100%: you always have to check the output carefully, and that also takes a lot of time. So I often think I could still do it better myself—and possibly faster—without AI.”

Differentiation becomes practical:

“Using AI helped me see ways to differentiate across different parts of a lesson, and it also provided concrete tools to do so. We all know we should differentiate, but the practical implementation is sometimes missing in advance.”

AI as preparation partner; next step is learner use:

“AI is certainly a valuable and practical tool for preparing (and designing) lessons, especially as a sparring partner that provides inspiration. It is also good that certain quality requirements are kept in view (e.g., applying Bloom’s taxonomy principles), which you can sometimes lose sight of when time is short. As a next step, it would be interesting to explore how AI can be used in class to support learners’ learning processes.”

Final comments

The GLOW project began in september of 2024. At that time, we had many teachers who felt unsafe, unsure and sceptical to the use of generative AI and its possibilities in the classroom. As a project group, Briga and KOV are convinced that the methodological use of generative AI in lesson planning is key for teacher development, and for teachers to find breathing room for themselves. The quotes above show a group that represents an eagerness to learn, thanks to this project. They also represent progress. Progress in their rediness to tackle new digital tools. Progress in prompting quality and progress in creating new materials and using new ideas, without the fear of failing.

The project aimed to lower the workload of teachers, and in some ways this project did indeed do that. In other ways, the workload increased, but in a manner that allowed for skills development or exploration. An unexpecte benefit which we welcome to the project.

The project set out to help teachers learn how to prompt, and to create prompts that would be able to be reused. In the chapters above, there are clear examples of prompts that have functioned, and there is a recipe which can be replicated. The goal has been achieved, and we hope that you, the reader, have learned something from this project.

The dedication of Angela, Barbara, Kuno, Murtaaz and Wouter cannot be understated, Briga and KOV hope that you find this useful and share it on with other users, who may one day benefit from it too.

Appendices

Appendix 1: Full prompts– selection for guide

Plans for the Christmas holidays

1) Role. You must be an experienced teacher of Norwegian as a second language who is good at creating teaching plans with good progression of high quality. You are used to teaching on zoom. And you have intercultural competence.

2) Theme. I would like you to create a lesson plan for a group at a weak B1 level. The theme is "plans for the Christmas holidays". We have an emphasis on oral skills in this session. Can you create a dialogue between two work colleagues? A Polish employee who travels to Poland during the Christmas holidays and a Norwegian employee who is visited by his in-laws. Both of them are employed by a cleaning company. They also talk a little about there being a Christmas party just before Christmas. The text should be approximately one A4 page.

Also make a list of important or difficult words in Polish, Lithuanian, Ukrainian, Thai and Albanian.

The participants will then read the dialogue aloud to each other with assigned roles in groups of two in breakout rooms on zoom.

We have worked with question words. Can you create questions for the text you have created and use question words in the questions? Participants must answer these questions orally.

We will also have a talking exercise on zoom in small groups. Can you create "interview questions" for participants to ask each other? The questions should be about Christmas celebrations, Christmas food, Christmas traditions or plans for the Christmas holidays. There should be at least 7 questions.

3) Time period: The session is 60 minutes, it takes place completely digitally on zoom. All participants have access to computers.

4) Target group:

The participants are at a weak B1 level. This is a more general Norwegian course, i.e. not a specific "working Norwegian course". There are about 10-12 participants in the group from Poland, Lithuania, Albania, Thailand and Ukraine. They are all women and are between 25 and 45 years old.

5) Come up with a voluntary homework. It can be writing an essay. Write that the homework is voluntary. No differentiation. The activities should be the same for everyone.

6)Tone: professional and didactic, but also warm and supportive.

7) Feel free to include suggestions for the teacher's instructions along the way. Come up with a short suggestion for a "warm-up", i.e. how the class can be started in a good way.

Feel free to write the teaching plan with headings and a good division/structure. The curriculum should be clear and easy to read.

8) The assessment consists of the teacher observing how the participants answer the questions with question words and whether they remember the question words. In addition, we use the last 5 minutes to summarize and repeat. Here, the participants will also tell if it is difficult to express themselves orally, and what it is that makes it difficult.

9) In order for the teacher to assess the participants' skills and for the participants to assess themselves, we need a box of Bloom's taxonomy. Can you create such a box when it comes to answering questions with question words and a box for the participants' conversation with interview questions on zoom?

10) Can you formulate an active learning strategy/active speaking strategies that can be of help to the participants when they need to practice speaking Norwegian?

Summer holidays

<prompt> <instruction> Make a lesson plan about "Summer Vacation" for course participants Norwegian as a second language. Think step by step, use if-then reasoning, and give brief pedagogical justifications. If there is a lack of information, make reasonable assumptions and mark them as "Assumptions". If something cannot be decided, write "I don't know" and suggest an alternative. </instruction> <context> Target group: adults, mainly B1 approaching B2. Frames: 12–18 participants, F2F, duration 90 minutes. Language focus: talk about experiences, ask/answer holiday questions, write short text. Supports: visual aids, model texts, sentence starters, word/phrase lists, pronunciation hints. Topics: travel, activities, weather, places (Norway/abroad), feelings/experiences. </context> <role> Experienced NfA teacher who briefly justifies why each activity supports the goals and target group. </role> <requirements> 1) At least 3 learning objectives formulated as SMART, marked with Bloom level and progression (B1→B2). 2) Active learning strategies with short justification (Think–Pair–Share, role-playing, jigsaw). 3) Full time distribution for the

whole session (90 min). 4) Enter material/resource type (without URL). </claim> <goals> 1. Title 2. Level and target group 3. Learning objective 4. Assessment criteria per objective 5. Core words and phrases 6. Grammatical focus areas 7. Materials and resources 8. Schedule (Heating · Input/Modelling · Listening · Conversation · Writing/Production · Conclusion) 9. Activities step-by-step (teacher instruction + student assignment) with methodology and justification 10. Differentiation by main activity (B1 support, B2 challenge) 11. Formative assessment (checklist/exit-ticket) + example sentences for feedback 12. Home assignment/continuation 13. Assumptions </template> <process> a) Goals → b) strategies → c) task sequence → d) differentiation → e) assessment. If predominantly B1: more modelling/shorter texts. If more close to B2: longer utterances, more connectors, role-playing with reflection. </process> <output> 1) Reader-friendly Markdown with headings/bulleted lists. 2) Short JSON "metadata": { "theme": "Summer vacation", "level": "B1–B2", "varighet_min": 90, "goal":[...], "core word":[...] }. Tone: friendly, clear. </output> <task> Produce the plan according to the 90-minute template. For each activity: (i) time estimate, (ii) instruction, (iii) justification, (iv) differentiation B1/B2, (v) assessment. </task> </prompt>

Summer holidays

Hello! Role: You are a Norwegian teacher for foreign language learners with 15 years of experience. Task: Now you will make a lesson plan for a teaching session in work-related Norwegian. Try to connect topics: summer holidays with their work (dialogues in the workplace, etc.) Parameters: It will be a physical teaching where the students sit in the classroom at their workplace. The session lasts 4 lessons and has a 15-minute break. Divide the teaching by duration for each activity. All participants have notebooks, and also smartphones that they can use. Some have a PC, but not all. The teacher has a PC, projector and blackboard. Target group: There is a group of 10 people at approx. A2 level. They work in the cleaning industry, cleaning both public buildings and private homes. They come from different countries from all over the world and are between 25-50 years old. Context: At this session, they will learn to talk about the summer holidays. There should be a warm-up, main part, summary and suggestions for voluntary homework (preferably some online assignments related to vocabulary). - They must become familiar with vocabulary, make a list of useful words related to the summer holidays and summer activities - They must repeat past tense forms to be able to tell about the past - the summer holidays. - Include several forms of active learning: - I would like them to describe some pictures that depict typical summer activities. Help me find pictures. - I want them to spend a lot of time on group work where the main focus is on oral exercises. They can create dialogues, or similar. - Create some

assignments, e.g. a longer hole text about the summer holidays that the students must fill in with appropriate verbs in the past tense. - I also want you to create quizzes / kahoots that help to remember new vocabulary and past tense forms of verbs they have used in class. - IMPORTANT: Make your proposal for the teaching plan in canvas so that I can make my remarks BEFORE you generate the teaching plan. Learning goal: I want them to learn the vocabulary related to the summer vacation, that they should be able to have a conversation about their vacation at their workplace, both with colleagues and superiors. And that they use the verbs in the past tense correctly when they talk about it. Style and tone: It should be a friendly but at the same time professional tone. I like to weave humor into the teaching and make people laugh. Extra: I will get a PDF document containing a list of words related to summer vacation, plus a list of at least 30 verbs (infinitive and past tense) that they can use to tell about it.

Car mechanic

You are a teacher of Norwegian with extensive experience in teaching adults from different backgrounds. You will start a new course with a group of 5-6 adults. The level is varied. They have conducted a mapping test, but you would like to get a better overview of the level of the participants to decide which resources/materials to use, and adapt the course plan. The goal of the course: they will function better in their everyday work, and communicate better in the workplace. They are car mechanics. Course content: mainly language courses, but the content should be practice-oriented and work-related. Create a course plan for an online course session. The lesson plan should be at least 1800 characters long (one A4 page). Duration of class: 2 hours. The lesson should consist of: 1. Warm-up: you can suggest a fun activity that makes the participants "wake up", and will create a good atmosphere to get a good framework for further work. 2. Some activities that can help determine at what level the participants really are. The skills to be tested today: a) listening comprehension, b) oral skills, c) simple grammar - e.g. verbs: infinitive, present tense and modal verbs. 3. A closing: summary, possibly another fun activity to end the session in a nice way.

Housekeeping hotels

You are an experienced teacher of Norwegian as a second language. Your expertise is improving communication skills.
Develop a qualitative lesson plan about the theme ‘Housekeeping Equipment Used in Hotels’ consisting the following mandatory elements:
· General Information (lesson title, target group, proficiency level, duration, class size and online context).

· Lesson Objectives (SMART formulated)
· Warm-up Activity
· Body of the lesson divided in different activities
The nice-to-have elements in the lesson plan are:
· Materials
· Assessment
· Reflection/Homework
The group of students consist of 40 adult students between 18 and 45 years old. Their goal is to get a jobs in Housekeeping department. They have some understanding of the language (A2 level) but can't use it to communicate effictively.

The lesson plan for the lesson number 1, modul "Equipment Used in Hotels". There are 4 lessons in total.
Lesson 1, Oral goal: to describe how to use equipment, how to inform the status of the quipment to the other departments.
Lesson 1, Grammar goal: Gradbøying av adjektiv
Specificity of learning objectives: Formulate the 3 learning objectives SMART with concrete examples and explicit connected to the subject-specific domains with clear progression indicators.
Taxonomic depth: Reference to Bloom's taxonomy to have detailed cognitive progression. Give clear level indicators and integrate with the subject-specific skills.
Active learning strategies: Suggest 3 active learning strategies that works well for the theme of this lesson. Also explain why you suggest these active learning strategies.
Make sure that differentiation on the cognitive level of the students is possible and explain also why you suggest this differentiation.
Contextual specification: Norwegian as a second language for Housekeeping Department in Hotels

Forsee in the lesson plan 3 listening fragments. The total duration of the lesson is 180 minutes. Online tuition. Students can't use their smartphone.

Tone is friendly but instructive, teacher-to-AI tone.

Self-presentation

1) Role. You must be an experienced teacher of Norwegian as a second language who is good at creating teaching plans with good progression of high quality.

2) Theme. I would like you to create a lesson plan for a group at A2-B1 level. The theme is "to present yourself". I am thinking here of two different ways of presenting oneself. 1)

A more informal way. How to present yourself to a new colleague during the lunch break? 2) A more formal way. How to present yourself during a job interview. I want you to create 2 texts. The first can often take the form of a dialogue. The other may be more like a monologue. In the second text (job interview), you present in the first person form a person who wants to get a job in a new company, and who has previous experience working for cleaning companies. The person must be from Poland, and not only introduces themselves with their name, age and professional experience, but also says something about hobbies and characteristics that define the person. The person also talks about their Norwegian language skills. For the texts, I want reflection questions for my course participants and I want a list of important/difficult words in the languages Polish, Lithuanian and Ukrainian. I also want the course participants to practice orally on both forms of presentation (informal and formal conversation) in group conversations in breakout rooms on zoom.

3) Time period: The session is 60 minutes, it takes place completely digitally on zoom. All participants have access to computers. There should not be a focus on grammar, but preferably a focus on idiomatic expressions with a preposition, e.g. to be responsible for something, to be concerned about something, to be good at something.

4) Target group:

By level A2-B1, I mean that the participants are at a good A2 level and are on their way to B1 level. You don't need to create texts at both A2 and B1 levels. The texts can all be at B1 level, but many of the words and expressions need to be explained and practiced. This is a more general Norwegian course. There are about 12 participants in the group.

5) 1. Yes, include all the elements of the GLOW structure. Feel free to write that the homework is voluntary. 2. No differentiation. The activities should be the same for everyone.

6- Tone: professional and didactic, but also warm and supportive.

Regarding step 7: Feel free to include suggestions for the teacher's instructions along the way. Leave the warm-up short.

Appendix 2: Checklist for ethical use of GenAI in the classroom

Ethical Use of Generative AI in the Classroom

The checklist below is based on the EU AI Act by the Future of Life Institute

1. Transparency and Disclosure

- Students clearly state when AI tools (e.g., ChatGPT, Copilot) were used in their work.
- AI-generated content is labelled or acknowledged.
- Teachers inform students when AI tools are used in teaching, grading, or feedback.
- Students understand that they may be interacting with AI systems.

2. Human Oversight

- Teachers review and verify AI-generated outputs before using them in instruction.
- AI tools do not make final grading or assessment decisions without human review.
- Students critically evaluate AI outputs instead of accepting them automatically.
- Teachers remain responsible for instructional decisions.

(The AI Act emphasises that AI systems must allow for human oversight and not replace human judgment in high-risk contexts such as education.)

3. Fairness and Bias Awareness

- AI outputs are checked for bias, stereotypes, or discriminatory language.
- Teachers discuss bias in AI with students as part of digital literacy.
- AI tools are not used to unfairly evaluate students' abilities or characteristics.

(AI used to evaluate learning outcomes or steer learning processes can be considered high-risk and therefore requires strong safeguards.)

4. Data Privacy and Protection

- Students do not upload personal or sensitive data to AI systems.
- Teachers avoid sharing identifiable student information with AI tools.
- Only approved AI tools that follow privacy policies are used in class.
- AI tools comply with data protection regulations.

5. Academic Integrity

- AI is used as a learning support tool, not a shortcut to complete assignments.

- Students understand the difference between AI assistance and AI plagiarism.
- Teachers provide clear guidelines on acceptable and unacceptable AI use.
- AI-generated work must be edited, verified, and meaningfully contributed to by the student.

6. Safe and Responsible Use

- AI is not used to manipulate, mislead, or harm others.
- Students avoid creating harmful or deceptive content (e.g., deepfakes).
- AI tools are used in ways that support learning and wellbeing.

(The AI Act explicitly prohibits manipulative AI systems and systems exploiting vulnerabilities such as age.)

7. Accuracy and Reliability

- Students fact-check AI outputs with credible sources.
- Teachers explain that AI systems can produce incorrect or fabricated information.
- AI-generated information is used as a starting point, not a final authority.

8. Responsible Educational Use

- AI supports learning objectives rather than replacing learning activities.
- AI tools encourage critical thinking, creativity, and reflection.
- Teachers evaluate whether AI improves or harms the learning process.

9. Digital Literacy and Awareness

- Students learn how AI works, including its limitations.
- Ethical and societal implications of AI are discussed in class.
- Students understand risks such as bias, misinformation, and overreliance on AI.

10. Accountability

- Teachers and institutions define clear policies for AI use.
- Violations of AI guidelines are addressed through academic integrity policies.
- Schools regularly review and update AI policies.

Quick classroom rule (student version)

Before submitting work that used AI, ask yourself:

- Did I tell my teacher I used AI?

- Did I check the accuracy of the information?
- Did I edit and contribute my own ideas?
- Did I avoid sharing personal data?
- Is this helping me learn, not replacing my thinking?

Appendix 3: Survey Templates of teachers

First survey – Before project start

Email
Name
Which school or organisation you work for?
E-mail
Your experience teaching
How many years have you been teaching?
How many years of experience do you have working with digital tools?
Which level of second language learning do you normally teach?
How often do you find cognitive skill differences impacting the pace of second language acquisition in your classroom?
How do you assess cognitive strengths and weaknesses in second language learners to tailor your lessons?
Do you use differentiated instruction to address varying cognitive levels?
If yes, how effective do you find it?
How often do you find cognitive skill differences impacting the pace of second language acquisition in your classroom?
What challenges do you face in adjusting your teaching methods to accommodate students with varying cognitive abilities?
To what extent do cognitive differences affect your use of learning strategies?
What kind of cognitive training or strategies do you employ to help slower learners improve in second language acquisition?
How effective are peer learning and group activities when dealing with students of varying cognitive abilities?
What technological tools or resources do you use to accommodate cognitive differences in your language learners?
What are the most common cognitive challenges faced by adult learners with limited educational backgrounds in your class?
How familiar are you with the concept of generative AI in general?
Have you ever used generative AI tools (e.g., ChatGPT, language models) in your second language teaching?
If yes, which version did you use and was it a free or paid version of the tool?
To what extent do you feel confident in using generative AI for lesson planning or classroom activities?
Please explain why you feel this way?
What challenges or concerns do you associate with the use of generative AI in language teaching?
To what extent do you think generative AI can reduce your workload in creating lesson plans or preparing learning materials?
Have you received any training or professional development on how to use AI tools in language education?
What was the focus of this training and what did you learn?

Final Survey

Your name
Your school/CVO
Your e-mail
How many years have you been teaching?
How many years of experience do you have working with digital tools in class?
What level of second language training do you normally teach?
Has the use of AI made changes for you in classroom differentiation - cognitive learner differences?
What changes do you assess most helpful?
Has the use of AI made changes for you in class differentiation - learning style (quick-slow learner / grammar focus – meaning focus learner / other)
What changes do you assess most helpful?
Do you as a teacher feel more confident and secure using AI (change January 2025 to January 2026)?
Do you as a teacher have more trust in AI (change January 2025 to January 2026)?
Did the use of AI reduce the workload (in both time and cognitive workload)?
Is AI inspiring as a sparring partner / critical friend / thinking assistent?
Do you recognise the human – machine – human way of working?
What are take-aways for you from this project? What did you learn? What do you want to share? (Maybe what will be next?)

Log – when handing in lesson plans

Your name
Who do you work for?
Lesson plan for which month?
Time spent on making the lesson plan from beginning to end (in minutes)
Classroom context
In which group are you participating?
If ChatGPT, which version?
If ChatGPT, did you write or read in your prompt?
Link to ChatGPT prompt with results (if not ChatGPT, write name of file)
Are there any challenges you're facing with prompting that we can help with? (optional)
Reminder: don't forget to email to your contactperson of the project two document: (1) your lessonplan and (2) the word-document with the prompt history and the result of the prompt (for those who use
How would you categorize the lesson (reading, writing, oral, gramma etc)?

Appendix 4: Methodology in the project

The GLOW project has used a scientific model to inspire its project activities, and this method has been used during the entire project period. There have been three distinct phases to the project.

Phase 1 has been a data collection stage, where data related to the teachers' backgrounds, their knowledge and their expectations have been checked. This was done to create a baseline, as well as creating a foundation upon which we could later check progress. Phase one established a control on all factors related to both the project aims, but also related to the wider impact goals which are important for reporting the Erasmus+ project.

Phase 2 involved creating an experiment. The experiment was to also follow the basic principles of a scientific experiment, where all groups would do the same thing, but various variables would be controlled. In the case of the project, teachers were assigned to one of three groups: Control, Free or Paid. Control was assigned to teachers who were to be handing in lesson plans without using ChatGPT or any other generative AI tools. Free was assigned to teachers who would be using the free version of ChatGPT available at the time, to create lesson plans. Paid was assigned to teachers who would use a paid version of ChatGPT to create lesson plans. Once split into groups, the teachers started to produce lesson plans. These plans were handed in every month (excl. July and August) and were also followed by the log questionnaire. This was done to allow comparability between time used by teachers in creating lesson plans. This phase resulted in data collection on a massive scale, with over 270 lesson plans created and analysed, as well as several hundred log answers to look through. Users using ChatGPT were also told to submit their prompts.

Phase 3 included analysing the prompts used and the quality of the lesson plans. A rubric was created to give a score to both the prompt and lesson plan. When all this data was collected, we were able to create quantitative analyses of how the use of ChatGPT in lesson planning reduced workload, as well as how teachers learned or developed in their prompting. The final aspect of this was to produce everything in a guidebook and present it on a website.

Simultaneously to Phase 2 and 3, the project has hosted three trainings on generative AI. Furthermore the partners have visited one another and worked on the project in Oslo and in Bruges, tracking progress and using the time together to observe the others organisation.

Appendix 5: Prompt assessment rubric

Criterion	What is Evaluated	Score Scale (0–3)	Reasoning Behind the Criterion
1. Specificity of Learning Objectives	Whether the prompt clearly defines what learners should achieve and how outcomes are structured.	0 = No objectives; 1 = Vague/general objectives; 2 = SMART objectives with categories (knowledge/skills/attitudes); 3 = Highly detailed SMART objectives with examples and domain links.	Clear objectives guide AI to produce **focused and measurable learning outcomes**. Without explicit objectives, AI tends to generate generic lesson plans lacking pedagogical direction.
2. Audience-Specific Parameters	How well the prompt describes the target learners (age, level, background, needs).	0 = No audience defined; 1 = Only general level mentioned; 2 = Includes level, age, background, group size; 3 = Detailed learner characteristics, prior knowledge, and context.	AI performs better when the **learner profile is clearly defined**. Detailed audience information ensures the lesson plan is **appropriate, relevant, and pedagogically aligned** with the learners' abilities and needs.
3. Taxonomic Depth	The extent to which the prompt encourages cognitive progression	0 = No cognitive levels; 1 = General mention of thinking levels; 2 = Explicit reference to Bloom's taxonomy; 3 = Structured progression across	Including cognitive frameworks helps ensure **deep learning rather than simple information transfer**, guiding

	(e.g., Bloom's taxonomy).	cognitive levels with clear indicators.	AI to design activities that move from basic understanding to higher-order thinking.
4. Complexity of the Prompt	Whether the prompt provides structured, multi-step instructions for AI.	0 = Very simple prompt; 1 = Some multi-step elements but unclear; 2 = Structured multi-step prompt; 3 = Highly structured prompt guiding detailed output.	Complex, structured prompts produce **more coherent and pedagogically complete outputs**. AI responds better when the prompt specifies steps, components, and expected structure.
5. Active Learning Strategies	The extent to which the prompt encourages interactive learning methods.	0 = Passive instruction only; 1 = Minimal activation; 2 = Several active methods; 3 = Clearly structured interactive strategies with varied engagement.	Active learning promotes **engagement, retention, and skill development**. Prompts that request interactive strategies lead AI to generate lessons that involve discussion, practice, and collaboration.
6. Effectiveness in Guiding AI	How well the prompt leads to high-quality	0 = Poor output (vague/disorganized); 1 = Partially relevant but	This criterion evaluates whether the prompt

	AI-generated lesson plans.	incomplete; 2 = Mostly good with minor issues; 3 = Highly structured and aligned output.	actually **produces useful and well-organized AI outputs**, which is the ultimate goal of prompt design.
7. Time Management Parameters	Whether the prompt includes timing and pacing instructions.	0 = No timing; 1 = Only total lesson duration; 2 = Time per section; 3 = Detailed timing with flexibility and pacing adjustments.	Time structure helps AI generate **realistic lesson plans** that teachers can implement directly, ensuring balanced pacing and transitions between activities.
8. Contextual Specification	The degree to which the prompt describes environment, resources, and teaching conditions.	0 = No context; 1 = Basic setting; 2 = Learning environment and materials mentioned; 3 = Detailed contextual and logistical factors.	Contextual details allow AI to create **practical and implementable lesson plans**, tailored to specific teaching environments (e.g., online, hybrid, classroom).

Appendix 6: Lesson plan assessment rubric

Lessonplan Evaluation Criteria

Criterion	What is Assessed (scale 1-5)	Reasoning Behind the Criterion
Structure & Clarity	Clear learning objectives, logical sequence of activities, and a coherent lesson flow.	A well-structured lesson helps students follow the learning process step-by-step and ensures that the teacher's intentions are transparent and measurable.
Alignment with Curriculum Standards	The extent to which the lesson connects to relevant curriculum goals and competencies.	Alignment ensures that the lesson contributes to required educational outcomes and that teaching remains purposeful and standards-based.
Engagement Strategies	Use of active learning methods such as discussion, collaboration, role-play, questioning, and differentiation.	Engagement strategies increase student participation, support different learning styles, and improve knowledge retention.
Clarity & Completeness	Inclusion of all required lesson elements and clear instructions for activities, materials, and timing.	Completeness ensures the lesson can be implemented smoothly and that no critical instructional components are missing.

Appendix 7: Lesson plan template (with example)

• Lesson Plan

- This is an example of a lesson plan. The green elements must be included in the teaching plans of the GLOW project. The yellow elements can be in the teaching curricula of the GLOW project.

1. General information

- **Lesson title**: Ordering food at a restaurant
- **Target language**: English
- **Skill Level**: A2 (Beginner to Lower Intermediate)
- **Duration**: 60 minutes
- **Class size**: 10-15 adult students
- **Context**: F2F, online or hybrid

• 2. Objectives of the lesson

- By the end of the lesson, students will be able to:
    - Use common phrases to order food and drinks (e.g. *"I want a coffee, please")*.
    - Understand the restaurant's vocabulary (menu items, prices, and polite expressions).
    - Practice conversational skills by interacting in a role-play scenario.

3. Materials

- Menu cards (visuals)
- Audio clips of people ordering food
- Awards with important restaurant expressions
- A blackboard and markers

4. Warm-up activity (5 minutes)

- **Activity**: Ask students, "When was the last time you went to a restaurant? What did you order?" Encourage couple discussions and sharing.

5. The main part of the lesson divided into different activities (xx minutes)

- **Example - Presentation of vocabulary and sentences (15 minutes)**

- **Activity**:
  - Introduce key vocabulary: menu items (e.g., "burger," "pasta"), polite phrases (e.g., "Can I have...?" "I want..."), and numbers (prices).
  - Play an audio clip of a customer ordering food, and have students listen for these phrases.

- **Example - Guided exercise (15 minutes)**

- **Activity**:
  - **Repetition exercises**: The teacher models sentences, and the students repeat.
  - **Controlled dialogue exercise**: In pairs, students use handouts to practice ordering food. Example of dialogue:
    - *Waitress*: "Hey, what do you want?"
    - *Customer*: "I want a burger, please."

- **Example – Oral activity (role play) (20 minutes)**

- **Activity**:
  - Divide the students into pairs: one plays the waiter and the other the customer.
  - Use the included menu card for a more realistic experience.
  - Students switch roles after a few minutes.
  - The teacher circulates to give feedback on pronunciation and accuracy.

6. Assessment (5 minutes)

- **Activity**: Invite a few pairs to role-play in front of the class.
- **Formative assessment**: Notice if students are using goal phrases correctly. Provide immediate constructive feedback.

7. Reflection and Homework (5 minutes)

- **Activity**: Invite students to share what they learned today and what they experienced during the role play.

- **Homework**: The pupils write a short dialogue (5-6 lines) about ordering food at a restaurant.