

Implementation of source-grounded Generative AI as support for flipped learning

Mario Soilán, Jesús Balado

CINTECX, Universidade de Vigo, Spain.

How to cite: Soilán, M.; Balado, J. (2026). Implementation of source-grounded Generative AI as support for flipped learning. In: 12th International Conference on Higher Education Advances (HEAd'26). Valencia, 15-18 June 2026. <https://doi.org/10.4995/HEAd26.2026.21569>

Abstract

This paper reports the implementation of a Generative AI (GenAI)-augmented flipped classroom in a fourth-year Topography and Construction course (bachelor's degree in Mechanical Engineering). Two student teams prepared and taught theoretical topics using source-grounded GenAI tools to synthesize instructor-curated sources and produce a report and presentation. After each peer-teaching session, students completed a 15-item quiz generated from the presenters' materials. Average scores were high and similar for presenters and the audience. A post-activity survey showed predominant use of NotebookLM as a GenAI tool and positive perceived usefulness for learning and text preparation, alongside moderate editing and only moderate trust without verification. The results support the feasibility of pairing source-bounded GenAI with peer teaching and immediate formative assessment, while reinforcing the need to teach verification and critical evaluation.

Keywords: *flipped classroom; generative AI; active learning*

1. Introduction

The flipped classroom (FC) is commonly characterized by shifting initial content acquisition to pre-class time and using in-class time for active, instructor-facilitated application and feedback (Abeysekera & Dawson, 2015). Early syntheses in engineering education describe FC as a family of implementations rather than a single method, with outcomes depending strongly on what replaces lecture time (Bishop & Verleger, 2013). Later reviews and meta-analyses have strengthened the evidence base: large-scale analyses in higher education report positive overall effects of flipped learning compared with lecture-based approaches across academic, intra-/interpersonal, and satisfaction outcomes, while also highlighting substantial heterogeneity and the importance of design moderators (Bredow et al., 2021).

A recurring conclusion in the literature is that “flipping” by itself is not a guarantee of improved learning; rather, benefits are associated with active learning structures and opportunities for

practice, feedback, and engagement during class. Controlled comparisons have shown that when active learning is held constant, flipped versus non-flipped sequencing can yield similar learning gains, suggesting the primary mechanism is active engagement (Jensen et al., 2015).

From a learning sciences perspective, FC aligns naturally with frameworks that prioritize constructive and interactive student activity. The ICAP framework, for instance, predicts increasing learning outcomes as engagement moves from passive to active to constructive to interactive (Chi & Wylie, 2014).

This work studies the integration of Generative Artificial Intelligence (GenAI) within FC designs in higher education. GenAI introduces a specific opportunity: rapid generation of formative practice materials (e.g., retrieval questions, worked explanations) from instructor-approved content. Importantly, flipped designs benefit when quizzes or checks for understanding are incorporated: meta-analytic evidence indicates quizzes can be a positive design moderator in flipped implementations (van Alten et al., 2019). Systematic reviews and syntheses describe emerging uses in tutoring, feedback, brainstorming, and learning autonomy (Qian, 2025) as well as concerns and risks: UNESCO calls for human governance, teacher training, and safeguards on privacy and responsible use in educational settings (UNESCO, n.d.).

A key pedagogical risk of GenAI is hallucinations: plausible sounding but incorrect content generated by the AI. One mitigation strategy is to constrain generation to provided sources and require citation to those sources. One way to do so is to use AI tools that allow users to upload documents and then generate study guides, briefing documents, and other outputs grounded in those sources, with citations. Table 1 shows candidate tools for source-bound formative assessment generation, where instructors can curate the input corpus and audit the outputs.

Table 1. Source-grounded GenAI tools considered for the activity

Tool	Owner	Features
NotebookLM	Google	Audio overviews (AI podcasts). Presentations, quizzes and infographics with source-based outputs. (<i>Google Notebook LM</i> , n.d.)
Microsoft 365 Copilot Notebooks	Microsoft	Native Office365 integration. Interaction with OneNote, Word, Excel (familiar ecosystem) (<i>Microsoft 365 Copilot Notebooks</i> , n.d.)
ChatGPT Custom GPTs	OpenAI	Pedagogical customization with strict instructions on how to guide students and pre-defined knowledge files. (<i>Introducing GPTs</i> , 2024)
Claude Projects	Anthropic	Large context window as project knowledge base, for deep thematic analysis (<i>Claude Projects</i> , n.d.)
Perplexity Spaces	Perplexity AI	Private files may be merged with web-based search for external verification or updates. (<i>Perplexity Spaces</i> , n.d.)

This paper reports results from the application of FC in a Topography and Construction course, in the fourth year of the bachelor's degree in Mechanical Engineering. The objectives of the paper are to answer the following Research Questions: (1) How do students use and perceive the usefulness, risks, and boundaries of GenAI when integrated into a flipped course design? (2) What patterns appear in post-topic quiz performance, and do presenter teams differ from audience peers? (3) What practical design recommendations emerge for responsible integration of source-grounded GenAI into FC courses in engineering domains?

The remainder of this paper is structured as follows: Section 2 shows the methodology and design of the study. Section 3 presents the results and discusses the research questions. Finally, Section 4 outlines the conclusions of this work.

2. Methods

This study was implemented in the first semester of the 2025/2026 academic year, in the undergraduate course on Topography and Construction. The GenAI-supported activity replaced two theory sessions (of 90 minutes each) with a flipped-classroom format.

Students (N=11) were split into two teams (5 and 6 members) to prepare and teach two theoretical topics: (i) Territory/land-use planning and (ii) Building Information Modelling (BIM) and energy efficiency in construction.

2.1. Activity design

The activity followed an *AI-augmented flipped classroom* approach: students learned by preparing to teach, using GenAI to accelerate document synthesis and organization, while keeping the *human-in-the-loop* through oral explanation and peer questioning.

The workflow per session was as follows:

Instructor guidance: The course instructor provided a list of GenAI tools and their features (Table 1) and, for each topic, a set of guiding questions and sources (course bibliography, web resources) that the topic should address.

Preparation of the topic: Each team was required to produce a written report synthesizing the topic by following the guiding questions, and an oral presentation with audiovisual support (e.g. using PowerPoint) based on the report. GenAI tools were allowed for every content generation step.

FC sessions: Each team delivered a 60-75-minute presentation, with 10-15 minutes allocated per student. Each team autonomously managed the division of work within its assigned topic.

Post-presentation assessment. After each presentation, the class completed a short quiz generated with NotebookLM using the presenting team's report as source material; results were

split for the presenting group (A) and the audience group (B). The quiz was announced in advance to encourage the audience group's attention.

Feedback survey. Students completed an anonymous survey (10 questions, multiple-choice, Likert and open-ended) capturing tool usage and perceptions of usefulness, prompting effort, editing, hallucinations, trust, and perceived learning among other factors (Table 2).

Table 2. Feedback survey items and response formats

Question	Type
Tool Usage. Which GenAI tools did you use for this activity?	Multiple choice
Usefulness. Rate the overall usefulness you perceive in GenAI for (a) text generation, (b) schemas/images/visualizations, (c) learning, (d) presentations	
Prompting. How much effort did you put into refining prompts to obtain a good result?	
Human-in-the-loop. In the content you generated and reviewed, how much did you have to edit the raw GenAI output?	
Hallucinations. Did you detect conceptual errors or factually wrong data in the GenAI answers?	Likert (1-5)
Trust. If you had to do this work without access to resources to verify the information, how much would you trust the GenAI answers?	
Learning. Compared with traditional methods (searching bibliography, writing reports) how do you rate your understanding of the topic after using GenAI?	
Self-confidence. If your classmates asked you a complex question about your topic, how confident would you feel answering it?	
Professional usefulness. What do you think about the use of GenAI tools in your future work?	
Pros and cons: What main advantages/disadvantages did you personally notice using GenAI for this activity? What risks and opportunities do you think it creates for learning and teaching?	Open-ended

2.2. Measures and analysis

Quiz performance: For each topic, quiz scores were analyzed on a 0–10 scale for Group A and B measuring means and standard deviations (SD).

Student feedback survey. Descriptive statistics were computed for the Likert 1-5 items and open-ended responses were summarized via lightweight thematic grouping.

3. Results and Discussion

3.1. Quiz performance assessment

Table 3 reports the quiz results, on a 0–10 scale, split by presenters (A) and audience (B). The quizzes were automatically generated by NotebookLM using the presenters' report and presentation as sources. Quizzes consisted of 15 single-choice questions with four possible answers. Overall, performance was high across both topics, and differences between both groups were small.

Table 3. Quiz scores by topic and role

Topic	Group A mean $\pm$ SD	Group B mean $\pm$ SD
Territory / Land-use planning	7.86 $\pm$ 0.31	7.64 $\pm$ 0.84
BIM & energy efficiency	8.72 $\pm$ 0.60	8.85 $\pm$ 0.83

The absolute scores suggest good immediate comprehension after peer teaching, and the lack of detectable differences suggests that, in this implementation, teaching the topic did not translate into higher immediate quiz performance than attending the topic, at least as captured by these quizzes and with the limitations of the small number of students.

3.2. Perception survey assessment

The survey showed that the most used tools were Gemini/NotebookLM (91.7%), followed by ChatGPT and Copilot (41.7%). Specific image generation tools were rarely used (8.3%).

Figure 1 shows the response distribution for Likert items. Figure 1a shows that students perceive GenAI most useful for text generation and learning support, as well as for presentations, while lower ratings were given to schemes, images, and visualizations. The prompting effort was relatively high (*average* = 3.6, Figure 1b), suggesting iteration rather than “one-shot” use. Similarly, output editing was moderate (*avg* = 3.1, Figure 1c), consistent with the “human-in-the-loop” concept where students retain responsibility for the final content. This is consistent with the answers collected regarding hallucinations (*avg* = 3.16; Figure 1d) and trust (*avg* = 2.92, Figure 1e), with students reporting mainly small errors especially when generating images that contain text; and moderate trust, requiring students to verify the content. Finally, students reported a better perceived learning with respect to their traditional study techniques (*avg* = 3.67, Figure 1f) and good self-confidence to answer questions about the material they prepared with GenAI (*avg* = 3.58, Figure 1g).

The survey ended with two open-ended questions. Regarding the professional usefulness of GenAI tools, the most frequent themes were efficiency, time saving and ease of access to

information (mentioned by 41.7% of answers). The overall sentiment was positive, with non-specific evaluations (e.g. ‘very useful’, ‘they are the future’). Regarding the main advantages and disadvantages perceived by the students, they highlighted time efficiency and information access as the main advantages, and overreliance and skill loss as the main disadvantages: students pointed out that reliance and quality of results are directly related with the prior knowledge on a given topic when working with GenAI tools.

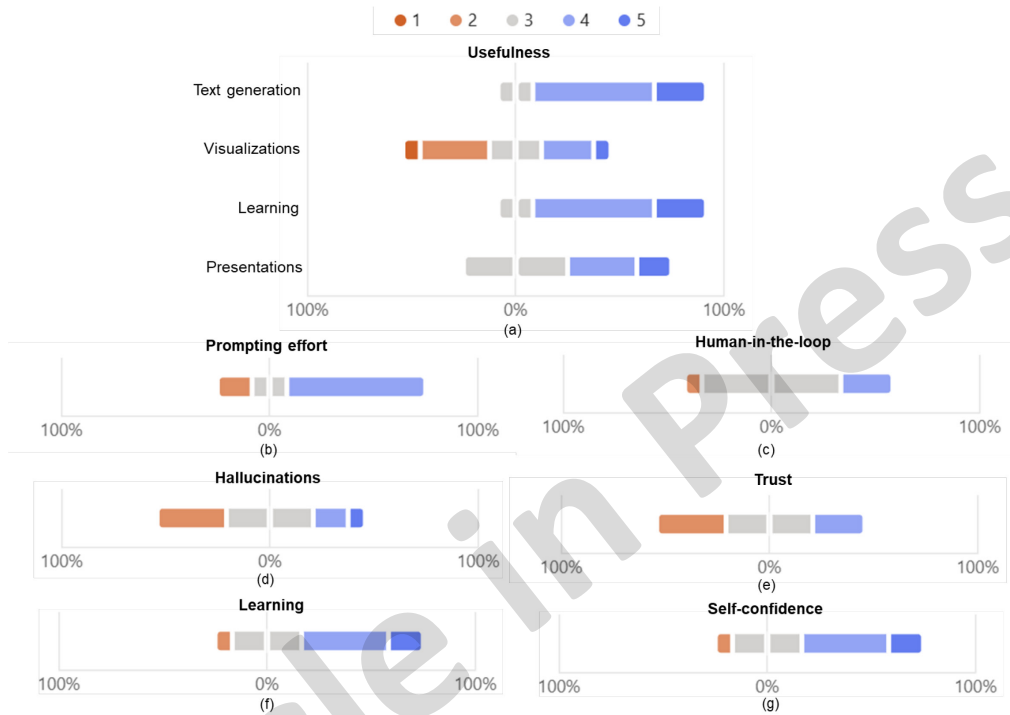


Figure 1. Distribution of responses to Likert-scale items (1-5) listed in Table 2

3.3. Discussion

The study supports the feasibility of pairing (i) source-bound synthesis through GenAI tools with (ii) peer teaching and (iii) immediate formative assessment. High quiz scores across both topics indicate that the approach can yield solid short-term comprehension, while the survey indicates students experience meaningful usefulness for learning and text-centric preparation without reporting high unconditional trust.

The instructor’s qualitative feedback was also positive. The quality of the oral presentations was good, the concepts were conveyed properly, and the attention of the audience peers was higher than in a standard session, in the instructor’s view. Students reported learning not only in the course topics but also how to use new GenAI tools, which is relevant as these tools are

constantly improving and offering new approaches to learning and generating content. However, this work also surfaces the fact that regulating GenAI use in higher education, although necessary, is highly complex as GenAI's capabilities seem to be improving faster than the regulatory capacity of a higher education organization, so it is important that instructors stay as up to date as possible with GenAI tools.

4. Conclusions

This work deployed a GenAI FC design in a Topography and Construction course, replacing two theory sessions with student-led teaching supported by GenAI tools. Key outcomes are:

- Short-term learning was effective, with relatively high quiz scores which did not differ meaningfully between presenters and the audience, suggesting that both roles can reach comparable immediate comprehension under this approach.
- NotebookLM was the preferred GenAI tool, indicating that students preferred multi-source grounded workflows with friendly user interfaces (e.g. a PDF presentation generated from source data only requires one click).
- Perceived usefulness was highest for learning and text preparation, with students reporting medium-high prompting effort and moderate editing, consistent with an intended human-in-the-loop model rather than full delegation.
- Students reported moderate trust without verification, and some detection of errors, reinforcing the need to explicitly teach verification strategies and critical reading.

For instructors considering source-bounded GenAI in flipped classrooms, the results point to a workable pattern: curate sources to work with and provide guiding questions as a starting point; require an accountable learning artifact (oral presentation of the learned material, post-presentation quizzes); and allow the use of AI for formative assessment.

The pilot is small-scale (two topics, small groups, short time window) and lacks a traditional control condition or pre/post measurement of retention. Future iterations should add (i) quality controls for generated quizzes, (ii) delayed tests to assess retention, and (iii) a rubric to document *how* GenAI was used by the students (prompting, verification actions) to better link process to learning outcomes.

Acknowledgments

This work has been supported by the Spanish Ministry of Science, Innovation and Universities through Grant RYC2021-033560-I funded by MCIN/AEI/10.13039/501100011033 and by European Union NextGenerationEU/PRTR

References

- Abeyssekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: Definition, rationale and a call for research. *Higher Education Research & Development*, 34(1), 1–14. <https://doi.org/10.1080/07294360.2014.934336>
- Bishop, J., & Verleger, M. A. (2013, June 23). *The Flipped Classroom: A Survey of the Research*. 2013 ASEE Annual Conference & Exposition. <https://doi.org/10.18260/1-2--22585>
- Bredow, C. A., Roehling, P. V., Knorp, A. J., & Sweet, A. M. (2021). To Flip or Not to Flip? A Meta-Analysis of the Efficacy of Flipped Learning in Higher Education. *Review of Educational Research*, 91(6), 878–918. <https://doi.org/10.3102/00346543211019122>
- Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Collaborate with Claude on Projects*. (n.d.). Retrieved 19 January 2026, from <https://www.anthropic.com/news/projects>
- Get started with Microsoft 365 Copilot Notebooks—Microsoft Support*. (n.d.). Retrieved 19 January 2026, from <https://support.microsoft.com/en-us/topic/get-started-with-microsoft-365-copilot-notebooks-0775e693-11c6-4d80-8aba-fcc81a737a06>
- Google Notebook LM*. (n.d.). Retrieved 19 January 2026, from <https://sites.google.com/view/notebook-lm>
- Introducing GPTs*. (2024, March 13). <https://openai.com/index/introducing-gpts/>
- Jensen, J. L., Kummer, T. A., & Godoy, P. D. d. M. (2015). Improvements from a Flipped Classroom May Simply Be the Fruits of Active Learning. *CBE—Life Sciences Education*, 14(1), ar5. <https://doi.org/10.1187/cbe.14-08-0129>
- Qian, Y. (2025). Pedagogical Applications of Generative AI in Higher Education: A Systematic Review of the Field. *TechTrends*, 69(5), 1105–1120. <https://doi.org/10.1007/s11528-025-01100-1>
- UNESCO: Governments must quickly regulate Generative AI in schools*. (n.d.). Retrieved 19 January 2026, from <https://www.unesco.org/en/articles/unesco-governments-must-quickly-regulate-generative-ai-schools>
- van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review*, 28, 100281. <https://doi.org/10.1016/j.edurev.2019.05.003>
- What are Spaces?* | *Perplexity Help Center*. (n.d.). Retrieved 19 January 2026, from <https://www.perplexity.ai/help-center/en/articles/10352961-what-are-spaces>