Teacher-Selected Sources, AI-Supported Outputs: NotebookLM in the CLIL Classroom

LUCAS, Craig

Abstract

Content and Language Integrated Learning (CLIL) requires materials that are both linguistically accessible and conceptually accurate, yet producing such resources is time-consuming and commercially available options remain limited. Recent attention has turned to generative AI (GenAI) tools such as ChatGPT, but their tendency to hallucinate, lack of source attribution, and potential for plagiarism limit their classroom reliability. This article examines Google's NotebookLM, a GenAI tool that grounds outputs in user-uploaded sources, offering verifiable traceability not available on most other platforms. Using open-access texts from *E-International Relations*, NotebookLM was employed to generate B1-level CLIL readings, comprehension questions, and discussion tasks for a Politics and International Relations CLIL seminar. Findings indicate that NotebookLM effectively reduces teacher workload while retaining disciplinary vocabulary and providing transparent source attribution. Generated materials were largely accurate, with comprehension questions that were mostly answerable and pedagogically usable. At the same time, outputs reflected the orientation of uploaded sources, underscoring that no text is neutral and highlighting opportunities to foster students' critical awareness of framing. Challenges include the need to update sources for currency, review simplification for conceptual precision, and ensure balance in perspectives. Overall, NotebookLM demonstrates strong potential to support efficient, reliable, and critically informed CLIL material development while keeping educators in control of content.

Keywords: *CLIL, Generative AI, Materials development, Critical literacy*

Introduction

A common concern for teachers of CLIL classes is the difficulty of finding or creating suitable materials. Commercially published CLIL materials are limited, and developing level-appropriate learning resources is highly time-consuming (Coyle et al., 2010). Recent research suggests that generative AI (GenAI) tools such as ChatGPT could help to address

this problem (Kohnke & Zou, 2025; Kostka & Toncelli, 2023; Wadden et al., 2024). However, several limitations have also been identified. These include a tendency to 'hallucinate,' or generate factually incorrect information (Bender et al., 2021; Dwivedi et al., 2023), the risk of unintentional plagiarism (Wadden et al., 2024), and issues related to bias, perpetuation of stereotypes, and the spread of harmful information (Dwivedi et al., 2023). Another major limitation is the inability of most GenAI platforms to cite or trace their sources (Kohnke et al., 2023; Kostka & Toncelli, 2023). Together, these issues hinder the production of accurate and reliable CLIL materials. The purpose of this article is to explore methods for generating factually accurate reading materials with identifiable sources by using Google's NotebookLM, a tool that offers source traceability not available in many other GenAI platforms. This article evaluates NotebookLM outputs in terms of source attribution, plagiarism risk, and classroom reliability, including the clarity of comprehension questions and the framing of issues.

Background

Content and Language Integrated Learning (CLIL) requires teachers to integrate content and language instruction. For example, in the Politics and International Relations class that I was in charge of in the spring semester, students discussed and presented about topics such as theories of international relations, war and peace, migration, and human rights. As suitable materials were difficult to find, I tried various ways of scaffolding and simplifying native-level materials, while retaining as much of the original meaning and concept as possible. One way to do this is by glossing less frequent or domain-specific words, as recommended by Nation (2022). This method is particularly useful in activities whose purpose is vocabulary development. For activities in which the goal is to comprehend and communicate ideas and opinions, however, simplified texts often proved more effective, as the learning burden was reduced, allowing students to focus on the content.

As text simplification is a labour-intensive process, many teachers and researchers are now exploring ways to use GenAI for support (Kohnke & Zou, 2025; Wadden et al., 2024). ChatGPT has proven useful for generating initial drafts and adjusting text levels, but its tendency to hallucinate and its lack of source attribution limit its classroom reliability, particularly in CLIL contexts where factual accuracy is essential. At the same time, although AI tools can save significant preparation time, they can also reproduce existing copyrighted material without attribution. This may pose a risk of plagiarism, since large language models draw from vast, undifferentiated datasets and cannot provide verifiable citations. These shortcomings highlight the need to explore other GenAI tools, such as NotebookLM, that may better address the challenges of accuracy and source transparency. The following section examines how NotebookLM can be applied in CLIL contexts.

NotebookLM is an AI-powered research and note-taking assistant from Google Labs that was publicly launched in 2023. Leveraging the Gemini model and Retrieval-Augmented

Generation (RAG), the free version enables users to upload and synthesize up to 50 sources, such as PDF files, web links, and YouTube videos. It offers grounded summaries, answers, and audio overviews in a podcast-like format (Martin & Johnson, 2023). A key advantage of NotebookLM over many other large language models is source traceability. Unlike models that generate responses from a broad, non-transparent training corpus, NotebookLM grounds its content in the documents uploaded by the user. The Gemini model supplies the grammar, vocabulary, and structure, while the factual claims are anchored in the chosen sources. This distinction is of particular relevance to CLIL contexts, where the quality of teaching materials depends not only on linguistic accessibility, but also on the accuracy and verifiability of subject-related content. Source-grounded output allows educators to design classroom tasks, such as expert group reading activities, comprehension checks, and analytical questions, while ensuring that learners can trace information back to identifiable sources. Such transparency aligns with the pedagogical goals of fostering critical engagement and academic integrity.

Methodology

Data Collection and Preparation

To prepare CLIL materials, I selected and uploaded open-access texts from *E-International Relations (E-IR)*, an independent non-profit publisher and website in the field of international relations. E-IR publishes articles and books aimed at students, academics, and general readers, making it an accessible source for classroom use. These materials were chosen for their relevance to course topics such as theories of international relations, war and peace, and human rights, and their open-access status ensured there were no copyright concerns.

Use of NotebookLM

The selected texts were uploaded into Google's NotebookLM (Google, 2023), which was then used to generate level-appropriate CLIL readings, comprehension questions, and discussion questions. Grounding outputs in pre-selected sources increased their accuracy, transparency, and reliability. This process reduced text preparation workload while maintaining informational accuracy.

Prompt Design and Refinement

Prompts were refined iteratively for consistency. The final version generated expert group reading sets (Dale & Tanner, 2012), in which small groups of students read a passage and answer comprehension and discussion questions, becoming 'experts' on that particular topic. Groups then disband and form new groups, where each student, as an 'expert' on the topic that they read about in their previous group, teaches their peers about that topic. The final prompt generated reading sets with the following five components:

1. A 200–250 word introductory text for all students to be covered as a class.
2. Three 150–200 word CEFR B1-level readings, each self-contained but contributing to an overall understanding of the theme. The purpose of using B1 is to keep the emphasis on content comprehension. More challenging readings for higher-level or highly motivated learners can be assigned as out-of-class readings. The prompt asks for all important domain-specific vocabulary to be retained.
3. Comprehension questions to check understanding, and discussion questions to encourage personal engagement and sharing of opinions with other group members.
4. A set of group discussion questions requiring students to synthesize and compare information across all readings when groups change.
5. A graphic organizer table summarizing main ideas, designed to support note-taking and collaborative synthesis.

Sample Output

The following excerpt illustrates the style, content, and accessibility of the generated readings. It is taken from an expert group's activity. The other two readings are entitled 'Socially Constructed Inequalities' and 'The Legacies of Colonialism and Systemic Inequality.' This reading was generated by NotebookLM from Bloor (2022). The embedded numbers (e.g., <1>) denote different parts of Bloor (2022) that are linked to in the original NotebookLM output. For reasons of space, only one sample reading is shown here. The full set of readings also includes synthesis questions and a graphic organizer, as outlined in the methodology.

Reading: Global Economic Inequality

The world today faces significant **economic inequality**, where wealth is not shared equally among countries or people <1>. This imbalance is often described by the **global north-south divide**: wealthier countries, mainly in the North (like Western Europe and North America), possess most of the world's income, despite having less than a quarter of the global population <1>. In contrast, less economically developed countries are primarily located in the South <2>.

To understand this, **world systems theory** is helpful <3> <4>. It divides the global economy into three parts: the **core**, the **semi-periphery**, and the **periphery <3>**. **Core** countries specialize in high-skill, capital-intensive production, controlling wages and manufacturing <3> <4>. **Peripheral** countries provide cheap labor and raw materials for the core's industries <3> <4> <5>. This system is considered "rigged in favor of the core"<3>, meaning it benefits richer countries and limits the development of poorer ones <4>.

Dependency theory also explains that poorer countries are kept underdeveloped to serve the needs of a global elite, or "global bourgeoisie"<3> <4> <6>. It suggests that the underdevelopment in the South results directly from policies and unfair trade practices by the North, a process sometimes called **neo-colonialism <7> <8> <9>**. Even international organizations like the IMF and World Bank are sometimes criticized for implementing Western-centric policies that worsen poverty and global inequality <9> <10>.

Key Terms: economic inequality, **global north-south divide**, **world systems theory**, **core**, **semi-periphery**, **periphery**, **dependency theory**, **neo-colonialism**.

Comprehension Questions:

1. What is the main characteristic of the **global north-south divide**?
2. According to **world systems theory**, what are the three main economic areas in the world?
3. How do **core** countries benefit from **peripheral** countries in the global economy?
4. What does **dependency theory** suggest is the reason for underdevelopment in the Global South?

Discussion Questions:

1. Do you think that the global economic system makes it hard for poorer countries to develop, as **world systems theory** claims? Why or why not?
2. What are some actions that could help reduce **economic inequality** between countries?

Findings and Discussion

Source Validation and Plagiarism Mitigation

One of the most significant advantages of NotebookLM is the ease with which sources can be traced. Each generated passage is anchored to notes or documents I uploaded, allowing me to confirm the factual basis of specific claims quickly. This transparency contrasts with ChatGPT, which produces fluent text but often leaves the reader uncertain about where particular facts, phrases, or examples originate. The traceability feature also helps mitigate plagiarism concerns. By reformulating content with source links, NotebookLM avoids the "uncannily similar" (Wadden et al., 2024, "HAL, are you okay?" section, para. 14) structures or wording observed by Wadden et al. (2024) in ChatGPT outputs. From a pedagogical perspective, this protects teachers from unintentionally using plagiarized material and provides an opportunity to model responsible source use for students. Materials created with NotebookLM can thus

demonstrate, in practice, how academic integrity is maintained when ideas are grounded in identifiable references.

Content and Comprehension Questions

NotebookLM's generated readings show a clear capacity to retain core disciplinary vocabulary while presenting abstract concepts in accessible form. For example, in the reading on global economic inequality, the text introduced frameworks such as world systems theory and dependency theory while maintaining key terms like *core, periphery,* and *neo-colonialism*. This suggests that the tool is less likely to produce content that is "generic and repetitive, lacking in nuance, and sometimes untruthful" (Kohnke & Zou, 2025), and instead can generate texts that preserve disciplinary depth. Unlike the "stochastic parrot" effect described by Bender et al. (2021), in which large language models reproduce patterns from opaque training data without grounding, NotebookLM anchors each claim in teacher-selected sources, offering verifiable traceability. This grounding enhances accuracy and also allows teachers to demonstrate critical source use in practice.

At the same time, the framing of generated readings reflects the orientation of the uploaded sources. In the inequality example, the emphasis fell heavily on structuralist critiques, including dependency theory and criticisms of Western-led institutions such as the IMF and World Bank. Such framing underscores the importance of recognizing that no text is neutral. Teachers can make this explicit by comparing differently sourced readings, encouraging students to analyze how perspective influences both content and interpretation.

Unlike the ChatGPT-generated comprehension questions with no clear answers observed by Wadden et al. (2024), NotebookLM produces mostly answerable comprehension questions, which is an important factor for classroom reliability. Most questions highlight key details or main ideas, although a minority are near-verbatim rephrasings of sentences in the text. With review and occasional editing, however, the questions can be made classroom-ready, reducing the need for teachers to rewrite large sets of items. This time-saving benefit allows greater attention to be directed toward higher-order discussion tasks and multimodal scaffolding, such as diagrams, maps, and illustrations. Taken together, the generated content and questions indicate that NotebookLM can produce materials that are both pedagogically usable and verifiable, while still requiring teacher oversight to ensure accuracy and critical awareness of framing.

Pedagogical Implications

Perhaps the clearest pedagogical benefit of NotebookLM is the way it reduces the burden of text simplification and question-writing, tasks that are essential in CLIL but also highly time-consuming. While ChatGPT can also generate simplified content, NotebookLM's advantage lies in its source-grounded output, which allows teachers to verify claims quickly rather than spending time fact-checking each detail. This reduction in preparation time

enables greater focus on higher-level teaching tasks, such as selecting information sources, designing activities, and adapting materials to learner needs. For example, more attention can be devoted to structuring and scaffolding 'expert group' activities (Dale & Tanner, 2012). NotebookLM's ability to generate multiple subtopic readings anchored to the same sources supports peer teaching and ensures that each student contributes unique information during the synthesis stage. It also creates space to integrate additional scaffolds such as photographs, maps, illustrations, and glosses. As Kohnke and Zou (2025) suggest, teachers can also spend more time customizing and differentiating learning materials by adjusting text complexity and designing culturally relevant examples. NotebookLM does not replace teacher expertise but redirects effort toward more interactive and differentiated lesson design. Because outputs are anchored in teacher-selected sources, educators retain control over the content, supporting both accuracy and alignment with course objectives.

Challenges/Limitations

Some challenges emerged when using NotebookLM to generate CLIL materials. First, the outputs are only as current as the sources uploaded. In fields such as international relations, where events develop rapidly, this can limit the immediacy of the content. One way to address this is to emphasize enduring concepts and theories in generated readings, which students can then apply to contemporary cases using their own background knowledge.

Second, simplification sometimes leads to imprecision. Complex terms such as *core/periphery* or *neo-colonialism* may be reduced to vague substitutes or oversimplified explanations, which risks obscuring their conceptual value. Teachers therefore need to review outputs carefully to ensure that key terminology is retained and accurately presented.

Finally, the framing of outputs reflects the orientation of the uploaded sources. Rather than being a defect, this can be pedagogically valuable: because NotebookLM makes its source base explicit, it allows teachers and students to recognize how perspective shapes content. Still, this places responsibility on educators to select a balanced range of sources and, where necessary, to supplement them with alternative perspectives.

Future Directions

An important direction for future research is how AI tools can be used to encourage critical thinking in CLIL classrooms. Rather than treating AI outputs as neutral, students could be guided to compare differently framed texts or to question the perspectives embedded in them. This connects to the need to examine how GenAI outputs may reflect theoretical or ideological biases. Prior studies of ChatGPT have noted tendencies toward particular framings, and in the CLIL context, such biases could shape both the comprehension and discussion tasks generated. Investigating how different tools privilege mainstream development narratives (e.g., colonial

history, trade imbalance, SDGs) or critical perspectives (e.g., world systems or dependency theory) would extend evaluation beyond accuracy and answerability to include the cultivation of critical literacy. A secondary area for research concerns how to balance accessibility with the retention of domain-specific vocabulary, since over-simplification risks distorting key concepts, while excessive technicality may overwhelm learners.

Conclusion

This study demonstrates that NotebookLM can play a valuable role in the development of CLIL materials by reducing the workload associated with text simplification and question design, while still allowing teachers to maintain full control over the content. Because outputs are grounded in teacher-selected sources, educators can ensure that readings and questions are accurate, verifiable, and aligned with course objectives. Rather than replacing teacher expertise, NotebookLM functions as an assistant that frees teachers to concentrate on higher-level pedagogical tasks such as designing interactive activities, integrating multimodal scaffolds, and differentiating materials to meet diverse learner needs.

At the same time, several practices can help maximize its effectiveness. Choosing a balanced set of sources helps avoid one-sided perspectives, while NotebookLM's explicit grounding makes framing more recognizable for both teachers and students. Retaining domain-specific vocabulary ensures that learners engage with key concepts while still benefiting from simplified syntax. Reviewing comprehension questions for redundancy or over-literal phrasing helps maintain their value as genuine checks of understanding.

Caveats remain: outputs are only as current as the sources provided, and simplification may lead to conceptual loss if left unchecked. Yet these limitations also highlight NotebookLM's pedagogical value: by making the influence of sources visible, it provides opportunities to foster students' critical awareness of how knowledge is framed. Overall, NotebookLM increases preparation efficiency and supports deeper engagement with both content and language, while keeping teachers firmly in control.

References

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? [Conference paper]. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (FAccT '21), 610–623. Virtual Event, Canada. https://doi.org/10.1145/3442188.3445922

Bloor, K. (2022). *Understanding global politics.* E-International Relations Publishing. https://www.e-ir.info

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge University Press.

Dale, L. & Tanner, R. (2012). *CLIL activities: A resource for subject and language teachers*. Cambridge University Press.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A.M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., . . . Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Google. (2023). *NotebookLM* [Large language model]. https://notebooklm.google

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, *54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Kohnke, L., & Zou, D. (2025). The role of ChatGPT in enhancing English teaching: A paradigm shift in lesson planning and instructional practices. *Educational Technology & Society*, *28*(3), 4–20.

Kostka, I., & Toncelli, R. (2023). Exploring applications of ChatGPT to English language teaching: Opportunities, challenges, and recommendations. *The Electronic Journal for English as a Second Language*, *27*(3). https://doi.org/10.55593/ej.27107int

Martin, R., & Johnson, S. (2023, July 12). NotebookLM: How to try Google's experimental AI-first notebook. *The Keyword*. https://blog.google/technology/ai/notebooklm-google-ai/

Nation, I. S. P. (2022). *Learning vocabulary in another language*. Cambridge University Press.

Wadden, P., Zardai, I. Z., Otsuru, S., & Umehara, H. (2024, February 16). *Creating and modulating course content using AI tools* [Paper presentation]. 5th International Symposium on Academic Writing and Critical Thinking, Nagoya University, Nagoya, Japan.