UNIT 10: Research Proposal Project Transcript

Long Term Effects of LLM Use on Independent Writing

Slide 1, Cover

(No voice-over)

Slide 2, Introduction

Large language models are now part of everyday academic writing, and they sit right at the intersection of learning support and academic integrity.

Research consistently shows that these tools can help students plan and refine their writing, improving fluency and organisation, with particular benefits for multilingual learners (Kasneci et al., 2023; Xu and Zhou, 2023).

At the same time, their use has blurred ideas of authorship and originality. Students often see AI assistance as acceptable support, while institutions tend to view the same practices far more cautiously. Chan's concept of *AI-giarism* captures this growing uncertainty around effort and ownership in student writing (Chan, 2025).

What remains unclear is the longer-term impact. While short-term improvements are well documented, there is far less evidence about what sustained use of LLMs means for students' ability to write independently over time (Almoraie and Alhejaili, 2025).

This study starts from that uncertainty and asks whether AI-supported writing strengthens or weakens independent academic skills in the longer term.

Slide 3, Review of the Literature, What the literature agrees on

Once we move past the headline debates, there is actually quite strong agreement in the literature on what LLMs currently do well as learning tools.
First, most studies report short-term improvements in writing quality, particularly in fluency, structure, and organisation. These gains are consistently observed across different contexts, but they tend to sit at the surface level of writing rather than in deeper skills like critical thinking or originality (Xu and Zhou, 2023; Tai et al., 2024; Almoraie and Alhejaili, 2025).
Second, there is broad agreement that LLMs can provide scalable formative feedback. While human feedback remains more nuanced and better prioritised, AI-generated feedback is often closely aligned with marking criteria and delivered instantly. In large

cohorts and resource-constrained systems, this makes LLMs attractive as complements to human feedback rather than replacements for it (Dai et al., 2023; Fan et al., 2024).

Third, the literature consistently highlights access and inclusion benefits. Multilingual and under-represented students, in particular, gain access to language support, explanations, and feedback that would otherwise require significant staff time or specialist provision (Kasneci et al., 2023; Ramírez-Montoya and Lugo-Ocando, 2024).

So at this level, the evidence is relatively stable. LLMs can support writing, scale feedback, and widen access. The unresolved question is not whether these benefits exist, but whether they translate into durable, independent academic skills over time.

Slide 4, Review of the Literature, Where the literature differs

While there is agreement on short-term benefits, the literature does diverges sharply on what those benefits actually mean.

The first disagreement concerns learning depth. Some studies report modest improvements in writing quality, while others find that gains remain limited to fluency and organisation. What is still unclear is whether sustained LLM use develops independent critical thinking or originality over time (Dai et al., 2023; Tai et al., 2024; Almoraie and Alhejaili, 2025).

Second, there is disagreement about how novel the integrity challenge really is. Chan (2025) frames AI-giarism as an extension of long-standing ambiguities around authorship, while other authors argue that the scale and invisibility of AI-generated text represent a more fundamental disruption to existing integrity frameworks (Bittle and El-Gayar, 2025; Zawacki-Richter and Marín, 2024).

Third, detection remains highly contested. Tools such as Turnitin’s AI writing indicator, GPTZero, and other commercial detectors are widely used, yet studies consistently highlight issues with false positives, false negatives, and uneven impact across student groups. This has led to disagreement over whether detection meaningfully supports integrity or simply introduces new risks, particularly for multilingual students (Susnjak and McIntosh, 2024; Pagaling et al., 2024).

Finally, the literature differs on governance and ethics education. While AI literacy is widely recommended, there is little agreement on which approaches actually change behaviour, and awareness of institutional policy alone does not reliably predict ethical AI use (Lund et al., 2026).

Together, these disagreements point to a lack of clear evidence about what actually supports learning and integrity over time, which is the gap this study addresses.

Slides 5 to 7, Review of the Literature, What's missing?

So what is missing from the current evidence base?

Most existing research focuses on short-term outcomes, early interventions, or student perceptions. There is very little work that tracks what happens to students' independent writing skills when AI use becomes sustained and routine.

This gap exists partly because these tools are still relatively new in higher education. As a result, we lack longitudinal evidence showing whether LLM-supported writing helps students develop critical thinking and originality, or whether those skills become dependent on the tool once it is embedded in everyday practice (Almoraie and Alhejaili, 2025).

This uncertainty matters because LLM use is no longer marginal. It is already shaping how students draft, revise, and submit work, often without consistent guidance.

In response to that gap, this study asks: *To what extent does sustained use of large language models for academic writing support the development of independent critical thinking and originality over time?*

Slides 8 and 9, Project Aims and Goals

The aim of this study is to move beyond short-term demonstrations of AI-assisted writing and examine how students' writing develops over time when AI is used regularly but not constantly.

Rather than focusing on performance with AI present, the study is designed to see what students can do on their own after repeated exposure to AI-supported writing. This places learning transfer and independence at the centre of the analysis.

To support that aim, the study has four practical goals.

First, it tracks changes in independent writing across the semester using repeated AI-restricted tasks.
Second, it compares those outcomes with performance on AI-supported tasks to separate assistance from learning.
Third, it evaluates higher-order writing skills using an analytic rubric that prioritises argument quality and evidence use, rather than surface fluency, which is where most AI gains occur (Dai et al., 2023).
Finally, it uses short reflective logs to understand how students use AI in practice, without relying on detection tools that are widely criticised as unreliable and inequitable (Bittle and El-Gayar, 2025).

Slide 10, Implementation

This study is implemented across a single academic semester to allow sustained exposure to AI-supported writing while keeping the learning context stable.

Students begin with a baseline AI-restricted writing task in Week 1. Across the semester, they then complete alternating AI-allowed and AI-restricted tasks, ending with a final AI-restricted task. This structure makes it possible to assess whether any improvement persists once AI support is removed.

The study will involve approximately forty to sixty undergraduate students. This sample is large enough to identify meaningful patterns over time, while remaining small enough for a single researcher to manage consistent marking, reflection analysis, and ethical oversight.

Slide 11, Writing and Reflection Tasks

This study uses a repeated, source-based writing task across the semester. Each task presents two short texts, one from peer-reviewed research and one from a policy or conceptual source, and asks students to evaluate the authors' claims and justify a position using evidence. The task format remains constant, while AI use alternates between allowed and restricted conditions.
This design makes it possible to test learning transfer. Improvement only when AI is available indicates assistance, whereas improvement in AI-restricted tasks provides stronger evidence of independent skill development, which remains a key gap in the literature (Almoraie and Alhejaili, 2025). The task focuses on higher-order academic skills such as argumentation and evidence-based judgement, rather than surface fluency alone (Howard, 1995; Dai et al., 2023).
After each task, students complete a short, AI-restricted reflection using Gibbs' reflective cycle. The structured stages support more systematic reflection than open prompts and are completed immediately after writing, while experiences and decision-making are still fresh (Gibbs, 1988). These reflections contextualise patterns of AI use over time without relying on detection or surveillance, aligning with developmental approaches to academic integrity (Bittle and El-Gayar, 2025; Chan, 2025).

Slide 12, Student Writing Task Evaluation

To evaluate these writing tasks, the study deliberately separates academic writing quality from language proficiency. This matters because research shows that AI support often improves fluency and surface accuracy without necessarily improving critical thinking or argumentation.
Academic writing quality is assessed using criteria informed by the British Academic Written English corpus, or BAWE. BAWE is a large collection of successful undergraduate writing from UK universities and reflects how academic writing is actually judged in higher education. Drawing on BAWE-informed marking practices, each task is assessed analytically for argument quality, use of evidence, critical engagement, and coherence, directly targeting independent reasoning and originality.
Language proficiency is assessed separately using IELTS and British Council writing band descriptors, covering clarity and coherence, lexical range, and grammatical control. Including these scores makes it possible to distinguish gains in language fluency, which are common with AI support, from deeper improvements in academic reasoning (Dai et al., 2023; Almoraie and Alhejaili, 2025).
Using both frameworks avoids treating polished language as evidence of learning and avoids reliance on AI detection tools, which are widely criticised as unreliable. Instead, the evaluation focuses on how students' academic writing develops over time, which sits at the core of the research question.

Slide 13, Study participation

Participation in this study is entirely voluntary. Students will be fully informed about the purpose of the research, what participation involves, and how their data will be used. Choosing not to take part has no academic consequences, and participation or non-participation has no impact on grades, progression, or assessment outcomes.

Slide 14, Data Collected

For students who consent, the study collects three types of data: their actual writing task submissions, the rubric-based evaluation scores, and short reflection logs completed immediately after each task. These submissions will be linked using a participation ID, but will not be used to identify the student who submitted them. They simply tie the chain of submissions to the same participant who submitted them allowing the tracking of the participants writing progress across the semester.

Slide 15, Data not collected

Importantly, the study deliberately avoids collecting unnecessary or intrusive data. No personal identifiers, system logs, keystroke data, browser histories, or AI detection outputs are gathered. This ensures the focus remains on learning outcomes and student experience, rather than surveillance or enforcement.

Slide 16, Ethical considerations

Ethical considerations in this study are addressed through specific, GDPR-aligned design choices rather than broad assurances. Participation is fully voluntary and based on informed consent, with students clearly told what data will be collected and, just as importantly, what will not. There is no academic penalty for participation or non-participation, and research data is kept entirely separate from formal assessment.
In line with UK GDPR principles of lawfulness, data minimisation, and purpose limitation, the study collects only three types of data: anonymised writing task submissions, numerical rubric scores, and short post-task reflective responses. No names, student ID numbers, demographic data, learning analytics, or AI interaction logs are collected.
This deliberate exclusion of identifying and behavioural data responds to concerns in the literature that surveillance-based approaches can undermine trust and disproportionately affect certain student groups (Pagaling et al., 2024; Chan, 2025). Participants retain full GDPR rights, including the right to withdraw and request deletion of their data at any stage without consequence.
The study also avoids AI detection tools. Existing research shows these tools are unreliable and can increase student anxiety while offering limited integrity value (Susnjak and McIntosh, 2024; Bretag and Mahmud, 2016). Instead, ethical AI use is supported through transparent task design and structured reflection, prioritising fairness, proportionality, and student wellbeing.

Slide 17, Artefacts created

Different artefacts will be created at different stages of the study. In the beginning, there will be the writing task papers that the students will complete, the consent forms that explain participation and data use, and the scoring sheets and rubrics that will be used to mark the work consistently.
During the project, most artefacts are generated by students. Writing tasks and reflective pieces can be completed either on paper or digitally, depending on student preference. Any handwritten submissions are scanned and transcribed into digital text, alongside their corresponding grading sheets. All materials are linked only by a

participant number, ensuring anonymity throughout.
After the project, two final outputs are created. The first is an online collection of anonymised writing tasks and reflections, labelled only by participant ID so progress can be tracked without identifying anyone. The second is a final summary report that brings together the results and sets out clear, evidence-based recommendations for how large language models can be used responsibly in academic writing.
Overall, this structure keeps the project organised, transparent, and ethically sound, while producing outputs that are useful beyond the study itself.

Slide 18, Conclusion

Large language models are no longer optional or experimental in higher education. They are already part of how students plan, draft, and refine their writing, and that reality is not going away.
Because writing is a lifelong skill, the real concern is not whether students use AI, but what that use does to their ability to think, argue, and write independently over time. If AI is left unexamined, it risks becoming a shortcut that weakens those skills rather than supports them.
This project matters because it shifts the conversation away from fear and detection, and toward evidence. By looking at writing development across a semester, with and without AI support, it allows us to see whether structured AI use strengthens learning or quietly replaces it.
The outcome of this study is not a blanket rule for or against AI. Instead, it aims to show when and how AI can be guided in ways that protect academic integrity while genuinely supporting learning. In that sense, the goal is balance, using data to inform practice, and ensuring that as AI becomes part of writing, it supports students rather than writing for them.

Slide 19, Thank you

Thank you for your time and attention.

Slide 20 - 21, References

Almoraie, A. and Alhejaili, M. (2025) 'A systematic literature review to implement large language models in education', *Education and Information Technologies*. https://doi.org/10.1007/s44217-025-00424-7

Bittle, K. and El-Gayar, O. (2025) 'Generative AI and academic integrity in higher education: A systematic review and research agenda', *Information*, 16(4), p. 296. https://doi.org/10.3390/info16040296

Bretag, T. and Mahmud, S. (2016) 'A conceptual framework for implementing exemplary academic integrity policy in Australian higher education', in Bretag, T. (ed.) *Handbook of Academic Integrity*. Singapore: Springer, pp. 463–480. https://doi.org/10.1007/978-981-287-098-8_24

Chan, C.K.Y. (2025) 'Students' perceptions of "AI-giarism": Investigating changes in understandings of academic misconduct', *Education and Information Technologies*, 30, pp. 8087–8108. https://doi.org/10.1007/s10639-024-13151-7

Dai, H., Zhu, X., Shen, S. and Li, Y. (2023) 'AI-generated feedback on writing: Insights into efficacy and ENL student writing', *International Journal of Educational Technology in Higher Education*, 20(1), p. 58. https://doi.org/10.1186/s41239-023-00425-2

Fan, X., Li, R., Zhang, Y. and Wang, Y. (2024) 'Comparing the quality of human and ChatGPT feedback on students' writing', *Journal of Writing Research*. https://doi.org/10.1016/j.jowr.2024.100021

Gibbs, G. (1988) *Learning by Doing: A Guide to Teaching and Learning Methods*. Oxford: Oxford Polytechnic.

Howard, R.M. (1995) 'Plagiarisms, authorships, and the academic death penalty', *College English*, 57(7), pp. 788–806. https://doi.org/10.2307/378403

Kasneci, E. et al. (2023) 'ChatGPT for good? On opportunities and challenges of large language models for education', *Learning and Individual Differences*, 104, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Lund, B.D. et al. (2026) 'Student perceptions of AI-assisted writing and academic integrity: Ethical concerns, academic misconduct, and use of generative AI in higher education', *AI in Education*, 1(1), p. 2. https://doi.org/10.3390/aieduc1010002

Pagaling, H., Cochrane, A. and Philp, J. (2024) 'Academic integrity or academic misconduct? Conceptual and practical tensions in higher education', *Higher Education Research and Development*, 43(8), pp. 1564–1584. https://doi.org/10.1080/07294360.2024.2339833

Ramírez-Montoya, M.S. and Lugo-Ocando, D. (2024) 'The impact of large language models on higher education', *Frontiers in Education*, 9, 1392091. https://doi.org/10.3389/feduc.2024.1392091

Shen, S. (2023) ‘Application of large language models in the field of education’, *International Journal of Emerging Technologies in Learning*, 18(4), pp. 24–35. https://doi.org/10.3991/ijet.v18i04.38017

Susnjak, T. and McIntosh, T.R. (2024) ‘ChatGPT: The end of online exam integrity?’, *Education Sciences*, 14(6), 656. https://doi.org/10.3390/educsci14060656

Tai, M. et al. (2024) ‘Impact of ChatGPT on ESL students’ academic writing skills’, *Smart Learning Environments*, 11, 2. https://doi.org/10.1186/s40561-024-00295-9

Xu, J. and Zhou, Y. (2023) ‘The impact of AI writing tools on the content and organisation of EFL student writing’, *Cogent Education*, 10(1), 2236469. https://doi.org/10.1080/2331186X.2023.2236469

Zawacki-Richter, O. and Marín, V.I. (2024) ‘ChatGPT in higher education: A synthesis of the literature and a bibliometric analysis’, *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-024-10334-1