

Policy Brief

LUMS Learning Institute | Scholarship of Teaching and Learning | 10 April 2026

Subject: AI-powered study tools in undergraduate writing assessment – evidence from a randomised controlled trial in SS-100 (Writing and Communication), Spring 2026.

Project title	<i>The Effect of AI-Powered Study Tools on Undergraduate Student Performance: Experimental Evidence from a First-Year Writing Course</i>
Research team	Faculty Leads: Dr. Ayesha Ali, Dr. Saqib Ilyas, Dr. Ihsan Ayyub Qazi, Dr. Agha Ali Raza. LLI Team: Laiba Intizar Ahmed, Rehana Kazi.
Reporting period	Spring 2026
Policy audience	VC Council; Dean's Council; OAA; Writing & Communication Programme; SS-100 Course Coordinators
Decision requested	<i>GenAI policy for early writing courses: (i) tool-free proctored assessments (the enforceable lever); (ii) institutional endorsement of source-grounded tools for course preparation; (iii) assessment redesign toward reflection-based and opinion-based writing; and (iv) embedded AI-literacy instruction.</i>

1. Executive Summary

Generative AI (GenAI) use among LUMS undergraduates is now near-universal: 85% report regular use and 60% use ChatGPT daily. Universities face a policy choice that no longer fits a binary frame; ban and fall behind, or allow and risk hollowing out learning. We ran a randomised controlled trial in a first-year Writing and Communication course (SS-100, Spring 2026), giving half the students access to NotebookLM (a source-grounded GenAI tool) during a proctored writing assessment and comparing performance to a no-AI control. Treatment improved overall scores by 5.6 percentage points ($p < 0.05$), driven almost entirely by higher-order synthesis tasks; on a personal-opinion writing item the treatment group performed worse than control presumably because the tool refused to generate the response and students had not read the passage closely enough. Realistically, LUMS cannot police which AI tools students use on personal devices when preparing at home. The enforceable institutional levers are therefore the assessment, the curriculum, and the tools LUMS officially endorses. Drawing on these findings, we offer the following considerations for the university's ongoing deliberation on GenAI policy. First, there is a case for restricting the use of GenAI tools in proctored assessments within foundational writing courses, where the development of independent writing capacities is central to learning outcomes. Second, source-grounded tools such as NotebookLM appear well-suited to structured preparation tasks than open-ended generative AI tools (e.g., ChatGPT, Gemini or Claude) and the university may consider providing institutional guidance and support for their adoption. Third, assessment redesign that foregrounds personal voice, reflective engagement and opinion-based engagement could help align evaluation with the skills least substitutable by generative AI tools.

Finally, embedding AI-literacy instruction across the curriculum would help students develop informed judgement about when to use, decline, or verify GenAI outputs.

2. The Institutional Problem

GenAI is now a default study aid for LUMS undergraduates. Our companion campus survey (N = 256, mean age 19.7 years, 38.5% female, freshman-to-senior) shows 85% use AI regularly. Reported uses are dominated by saving time (84%), brainstorming (81%), and writing/grammar help (79%). At the same time, 88% are concerned about over-reliance, 87% about reduced critical thinking, and 84% about ethical risks; 45% are extremely concerned about being falsely accused of misuse, and confidence in AI-detection tools is low.

LUMS lacks a coherent, evidence-based policy on GenAI use in coursework, particularly in foundational courses where writing skills are first developed. A blanket ban risks placing LUMS graduates at a competitive disadvantage; an unconditional permission risks weakening the very skills these courses are designed to build. Most existing evidence on AI and learning is observational and confounded by self-selection. SS-100 is the most consequential course in this regard: it provides the first sustained writing instruction most undergraduates receive, and the instruction most easily circumvented by AI tools.

3. What We Studied

We conducted a randomised controlled trial integrated into the Spring 2026 SS-100 course across 12 sections. Students were pre-assigned 50/50 (stratified within section) to a treatment arm with NotebookLM access or to a no-AI control arm, and completed a proctored 60-minute writing assessment worth ~3% of the course grade. Treatment students received a brief standardised tutorial (text + ~2-minute video, with a minimum-viewing-time control) before beginning the task. NotebookLM was selected because it is source-grounded, it answers only from materials the student uploads, which lowers hallucination risk relative to general chatbots and keeps engagement on course content. All other AI tools were blocked in both labs.

The assessment used a ~1,400-word reading passage and a common rubric covering (a) *lower-order skills*: identifying terms, interpreting tone, personal reflection and (b) *higher-order skills*: identifying claims and evidence, evaluating arguments, and writing a mini-op-ed. Section instructors graded responses blind to treatment status. All regressions include section fixed effects and baseline covariates; randomisation produced balance on 8 of 9 covariates. We complemented the regression analysis with sentence-transformer similarity scoring between student responses and canonical NotebookLM outputs, to detect cognitive offloading.

4. Key Findings

- **Overall scores rose by 5.6 percentage points** in the treatment group ($p < 0.05$). The gain is concentrated in the higher-order synthesis task (Task 2: +9 pp); the simpler comprehension task (Task 1) shows no overall effect.

- **The largest gains went to students with the lowest baselines.** Non-STEM students gained ~11.6 pp on overall score versus ~1.8 pp for STEM students (consistent with a ceiling effect). Female students gained more than males, narrowing a baseline gender gap on Task 2.
- **A specific writing sub-task moved in the opposite direction.** On the personal-opinion / personal-reflection item (L3), treatment students scored lower than control. This is the only item with a suggestive negative effect.
- **Cognitive offloading is visible in the text.** Sentence-transformer similarity analysis of treatment responses on Task 1 Part 3 shows near-verbatim reproduction of the canonical NotebookLM response structure; strong evidence that students were copying rather than composing. Such cognitive offloading in foundational courses that are designed to build mastery of core competencies, such as critical reading and analytical writing, can result in “never-skilling,” where students bypass the sustained practice needed to develop those abilities.
- **The mechanism on the personal-opinion item is informative.** NotebookLM, being source-grounded and designed not to express personal opinions, refused to produce the answer. Students who had been offloading earlier tasks to the tool had not read the passage closely enough to write a strong personal response in the time available.
- **Effects reflect tool use, not durable learning.** The 5.6 pp gain measures performance during AI access, not retained knowledge. The copy-paste similarity pattern cautions against reading the gain as a learning signal.

5. What This Means for LUMS – Policy Implications

(a) The headline gain is real, but it conceals a fragility

Treatment students did better overall presumably because they had access to a GenAI tool. But the moment the tool refused, on the personal-opinion item, their performance dropped below control. The apparent advantage is shallow: when the AI cannot answer, students who relied on it were unable to fully recover. This is the most important institutional finding in the study, and it should anchor LUMS's policy stance.

(b) The proctored assessment is the enforceable lever, and it should be tool-free in early writing courses

Foundational writing courses exist to build the very skills GenAI most easily substitutes for: close reading, voice, argument, and synthesis. Permitting AI during the assessment converts the assessment into a tool-use evaluation rather than a writing evaluation. The proctored lab is also the one setting where LUMS can actually enforce restrictions, by blocking tools at the network level (as we did in this study). The graded assessment in SS-100 and similar early-stage writing courses should therefore be tool-free.

(c) For at-home preparation, LUMS should endorse source-grounded tools but should not pretend to police private use

Students will continue to use ungrounded chatbots (ChatGPT, Gemini) on personal devices when preparing at home; no campus policy can credibly prevent this. The honest institutional play is to compete on quality rather than to prohibit. LUMS should officially endorse source-grounded tools (e.g., NotebookLM) and pre-load them with course materials. The case rests on closed-book design: grounded tools answer only from instructor-curated sources, lower hallucination risk, and keep engagement on course content; ungrounded chatbots can do the conceptual work for the student. Pair endorsement with explicit guidance, taught, not just stated, on when grounded tools help and when ungrounded ones substitute for learning. This is a recommendation backed by pedagogical reasoning, not an enforceable ban on private use.

(d) Assessment design must privilege personal reflection and opinion-based items

Because we cannot control which AI tool a student uses at home, assessment design, not tool prohibition, is the most durable defence against superficial AI gains. Tasks that demand personal reflection, lived experience, or value judgement are the items current AI tools are least able to short-circuit and, as we found, are likely to refuse outright. Re-weighting writing assessments toward such items preserves assessment validity in an AI-saturated environment and encourages the deeper engagement with source material that GenAI tools tend to bypass. This recommendation does double duty: it is the answer both to in-room AI use (when permitted) and to undisclosed at-home AI use.

(e) Equity dimension: targeted, supervised use can narrow gaps

Non-STEM and lower-baseline students gained the most from access. Permitting source-grounded AI in preparation (not assessment) for these populations could be an equity lever, provided the tool is grounded, instructor-scaffolded, and accompanied by explicit guidance on when not to rely on it.

(f) Beware the cognitive-offloading default

Near-verbatim reproduction of AI outputs in our textual analysis shows students will likely, by default, copy. Faculty should treat copy-paste similarity from grounded tools as a signal of disengagement, even when the response is technically correct. Rubrics should reward originality of synthesis and use of student-specific evidence.

(g) AI literacy is now a foundational skill and is the right substitute for unenforceable bans

Students need to be taught when to use AI, when to refuse it, how to verify its outputs, and why the choice between a grounded and an ungrounded tool matters for what they will actually learn. This is a new component of the writing curriculum, not an optional add-on. The personal-opinion failure observed in our study is a pedagogical opportunity: it is a concrete case where the right answer was to read the passage more carefully and write from one's own perspective. Embedding such cases in instruction does the work that a prohibition on private tool use cannot.

6. Recommendations for Action

#	Recommendation	Lead owner	Timeline	Resources	Risk if not acted upon
1	Prohibit GenAI tools (grounded or ungrounded) during all proctored assessments in SS-100 and other foundational writing courses	Writing & Communication Programme; SS-100 Coordinators	Fall 2026	Low	Assessment validity erodes; writing skill development undermined.
2	Officially endorse a source-grounded tool (e.g., NotebookLM) into SS-100 and possibly pre-load course materials. Frame as recommendation, not enforcement, for at-home use.	SS-100 Faculty; LLI	Next semester	Low-Medium	Equity gap widens; students default to ungrounded chatbots without guidance.
3	Redesign at least one major writing item per course to centre personal opinion, lived experience, or value judgement. Treat this as the primary defence against undisclosed at-home AI use.	Writing Programme; OAA	Next semester	Low	AI-tool gaming continues to inflate scores; voice and authenticity decline.
4	Issue a campus policy that (a) prohibits all GenAI in proctored assessments without explicit faculty authorisation, (b) endorses source-grounded tools for course preparation, and (c) requires student disclosure of AI use in non-proctored work.	VC Council; OAA	1 year	Medium	Inconsistent rules across schools; integrity disputes; student vulnerability to false accusations.
5	Embed an AI-literacy module, when to use, when to refuse, how to verify, and why grounded vs. ungrounded matters, into all first-year courses with substantial writing.	LLI; First-Year Programme	Next semester	Medium	Students continue cognitive offloading without judgement; private tool use stays unexamined.
6	Replicate the RCT in (1) other foundational courses (math, sciences, methods) and (2) advanced courses.	Faculty Leads; LLI	Strategic (1 year)	Medium	Policy generalised on partial evidence.

7. Implementation Considerations

- **Cost and staffing:** Recommendations 1-3 are low-cost; they are policy and assessment-design choices. Recommendations 4-6 require LLI capacity, faculty workshops, and clarity on institutional licence terms (NotebookLM is currently free; this may change).
- **Governance owner:** VC Council to issue the policy framework; OAA and the Writing & Communication Programme for SS-100 specifics; LLI for evaluation, faculty development, and replication studies.

- **Dependencies:** A clear policy definition of “source-grounded”; updated rubrics that reward personal voice; sufficient proctoring capacity for tool-restricted lab assessments.
- **Scope and enforceability:** Policy applies to LUMS-controlled settings: proctored assessments, course design, institutionally provided tools, and required activities. It does not, and cannot credibly, regulate which AI tools students use on personal devices when preparing at home. The brief is therefore explicit about what is enforced (assessments), what is provided (grounded tools), and what is taught (AI literacy and assessment-resistant tasks).
- **Risks:** Faculty resistance to assessment redesign; perception that LUMS is “behind” if any restriction is read as anti-AI; uneven enforcement across sections; students migrating to undisclosed external tools at home, which assessment redesign and AI-literacy must absorb.
- **Indicators of success:** Stable or improved scores on tool-free authentic writing tasks; reduced LLM-similarity scores in submitted writing; closing of the non-STEM/STEM and gender gaps under the new policy; positive student feedback on AI-literacy modules; replication in other courses confirming task-specific patterns.