Bond University

Legal Education Review

Volume 36 Issue 1

2026

Student Experience of AI in Law Assessment: An Exploratory Study

Benedict Sheehy
Faculty of BGL, University of Canberra

Juan Diaz-Granados
Thomas More Law School, Australian Catholic University

Follow this and additional works at: https://ler.scholasticahq.com/


© Copyright the authors.
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivative 4.0 International Licence.

STUDENT EXPERIENCE OF AI IN LAW ASSESSMENT:

AN EXPLORATORY STUDY

BENEDICT SHEEHY,[*] JUAN DIAZ-GRANADOS[◊]

ABSTRACT:

Legal education faces a growing challenge in responding to Generative Artificial Intelligence ('GenAI'). Students will encounter these tools in study and in practice, yet uncritical reliance on them risks undermining the development of legal reasoning, doctrinal accuracy and professional judgment. This article reports an exploratory study of a first-year Tort Law assessment in which students were required to generate a GenAI draft of a Letter of Advice, critique its adequacy, and then produce their own legal work. Drawing on in-class polling, class discussion and written student reflections, the article finds that students generally valued the exercise while remaining sceptical of GenAI's capacity to produce reliable legal analysis. The pedagogical value of the intervention lay less in AI-assisted drafting than in the development of evaluative judgment through comparison and critique. The article argues that a structured use-and-critique model offers a promising way to integrate GenAI into legal assessment without displacing foundational legal reasoning and writing skills. It does not claim generalisable findings. Rather, it offers a bounded account of one assessment intervention and its implications for assessment design and GenAI literacy in legal education.

I INTRODUCTION

Generative artificial intelligence ('GenAI')[1] poses a particular challenge for legal education. Unlike earlier educational technologies,

[*] Faculty of BGL, University of Canberra

[◊] Thomas More Law School, Australian Catholic University

[1] For a taxonomy of artificial intelligence ('AI') systems, including GenAI, see Juan Diaz-Granados, *AI and Tort Liability: Rethinking, Recalibrating, and Reallocating Risk and Responsibility* (Routledge, 2026), ch 2, s C. According to this author, GenAI 'involves techniques that enable the creation of new content—such as text, images, audio and video—derived from existing data. It includes tools such as generative

GenAI systems can produce coherent and professionally styled legal prose with remarkable speed. Yet that fluency masks persistent weaknesses: fabricated authorities, shallow doctrinal reasoning, poor factual application and a failure to appreciate the ethical and professional dimensions of legal work. For novice law students, who are still developing the intellectual frameworks needed to distinguish persuasive language from competent legal analysis, those weaknesses are not always immediately visible.

The central problem for legal education is therefore not simply whether GenAI should be permitted or prohibited. A purely prohibitory approach may reduce immediate integrity risks, but it offers little preparation for a professional environment in which GenAI tools are increasingly present. Conversely, unstructured adoption risks encouraging premature cognitive offloading before students have developed the reasoning skills, authority sensitivity and professional judgment that responsible use requires.[2] The pedagogical question is not whether students can produce legal text with GenAI, but whether legal education can design learning experiences that strengthen students' capacity to *judge* the adequacy of such output against legal and professional standards. [3]

Assessment is a critical site at which that question must be addressed. Assessment is where disciplinary standards, student strategy and professional expectations converge, and it is where the temptations and risks of GenAI become most concrete. If legal education is to respond constructively to GenAI, assessment design provides one of the most direct and influential levers available.

Courts have already underscored the relevance of this issue for professional practice. In JNE24 v Minister for *Immigration and Citizenship*,[4] Gerrard J observed that while there is 'nothing inherently impermissible' about using generative AI to assist legal research, it is 'not an appropriate substitute' for this legal skill and carries significant risks if used uncritically.[5] Such judicial warnings matter for legal education not merely as cautionary tales, but because they highlight the precise form of judgment students must learn: the ability to recognise fluent but inadequate legal work and to take responsibility for verifying, applying and communicating the law.[6]

---

adversarial networks and LLMs [large language models], like ChatGPT, Gemini and Llama': at 12.

2 Michael Veale et al observe that cognitive offloading occurs when students rely on AI to think and produce points and written structures: Michael Veale et al, 'Artificial Intelligence, Education and Assessment at UCL Laws: Current Thinking and Next Steps for the UK Legal Education Sector.' (UCL Legal Studies Research Paper Series 04/2025, 2025) 1, 12.

3 A question that must be addressed alongside the key issue of what to teach in law school: see Benedict Sheehy, Juan Diaz-Granados and Tomas Fitzgerald, 'What To Teach When Teaching Law: The Categories, Rights, Duties And Test ('CRDT') Framework' (2022) 19(1) *Canberra Law Review* 36.

4 *JNE24 v Minister for Immigration & Citizenship* [2025] FedCFamC2G 1314, [20].

5 Ibid.

6 Ying Chen and Benedict Sheehy, 'Educating Future Lawyers For Co-Intelligence: A Proposal for a Third Pillar in Legal Education based on a Comparative Study of the

This article approaches the pedagogical challenge posed by GenAI through assessment design. It reports an exploratory study of a first-year Torts assessment in which students were required to (1) generate a GenAI draft of a Letter of Advice, (2) critically evaluate that draft, and (3) produce their own legal advice. The purpose of the task was not to test whether GenAI could generate acceptable legal analysis. Rather, it was to require students to confront GenAI output directly and to assess its adequacy by reference to doctrinal accuracy, reasoning quality, factual application, audience awareness and professional responsibility. The deliberate sequencing of the assessment allowed the subsequent legal writing exercise to build on the prior work. In this design, GenAI functioned not as a surrogate author, but as an object of critique.

This article contributes to emerging scholarship on GenAI in legal education by examining how GenAI may be used, not as a drafting aid, but as a pedagogical comparator through which students can develop evaluative judgment, doctrinal accuracy, factual application, legal reasoning and professional communication. Its central claim is that the educational value of GenAI in early legal education lies less in efficiency gains and more in its capacity to make the standards of competent legal work visible through structured comparison and critique. Thus, it is not aimed at opposing the outdated prohibition on its use but addressing the realistic concerns of ubiquitous GenAI tools.[7] Pedagogically, if legal education is to respond constructively to GenAI, assessment design is one of the most important sites in which that response must occur. [8]

The study is deliberately exploratory and pedagogically oriented. It does not seek to produce statistically generalisable findings or causal claims about the effects of GenAI on learning outcomes. That limitation does not weaken the argument advanced here. On the contrary, it reflects the appropriate methodological posture for an emerging area in which legal educators are still experimenting with assessment design. The aim is to offer a carefully bounded account of one intervention and to draw out its implications for legal assessment and GenAI literacy, particularly in the formative stages of the law degree.

The article proceeds as follows. Part II situates the study within current debates about GenAI, legal education and the central role of assessment. Part III explains the assessment design and its pedagogical rationale. Part IV outlines the method and data sources. Part V presents the student feedback and thematic findings. Part VI analyses those findings in terms of evaluative judgment, experiential learning and professional formation. Part VII identifies limitations and implications for future assessment design. Part VIII concludes by arguing that

US and Australia Approaches' (2026) 43(3) *Wisconsin International Law Journal*, 243.

7 See Russell Bell, 'Between Prohibition and Practice: Institutional AI Policy Contradictions' (Working Paper, SSRN Electronic Journal, 2025).

8 See Qi Xia et al, 'A Scoping Review on How Generative Artificial Intelligence Transforms Assessment in Higher Education' (2024) 21(1) *International Journal of Educational Technology in Higher Education* 1.

supervised, structured and critical engagement with GenAI offers an educationally productive response.

The argument is not for uniform integration across all stages of the law degree. A staged approach remains preferable: foundational skills should be developed before students are encouraged to rely more extensively on GenAI, and early exposure should occur under supervision and with explicit emphasis on critique.[9] The present study is consistent with that position. It introduces first-year students to GenAI early, but within a controlled assessment design intended to reveal the limits of the tool as well as its attractions.

II GENAI, LEGAL EDUCATION AND WHY ASSESSMENT MATTERS

GenAI has intensified an existing question in legal education: how should law schools prepare students for professional environments shaped by new technologies without weakening the development of foundational legal skills?[10] That question is especially acute in law because competent legal work cannot be reduced to the production of plausible text.[11] Legal analysis requires authority-sensitive reasoning, identification of legally relevant facts, discrimination between stronger and weaker arguments, audience-sensitive communication and judgment about risk, uncertainty and professional responsibility. The pedagogical problem is therefore not simply whether students should encounter GenAI, but how that encounter should be structured so that it develops intellectual capacity and professional judgment rather than dependency.[12]

The literature reflects no single settled position.[13] Views on GenAI in higher education and legal education range from alarmist resistance to enthusiastic adoption.[14] Some scholars emphasise its potential benefits for assessment and learning, including rapid feedback, idea

---

9 See Chen and Sheehy (n 6); Luke Repici, 'Learning by Doing: The Importance of Limited and Thoughtful Generative AI Use in the 1L Legal Writing Classroom' (2025) 9(3) *Stetson Law Review Forum*, 1.

10 Chen and Sheehy (n 6).

11 Stefan Koos and Sebastian Wachsmann, 'Navigating the Impact of ChatGPT/GPT4 on Legal Academic Examinations: Challenges, Opportunities and Recommendations' (2023) 6(2) *Media Iuris* 255; Chunpeng Zhai, Santoso Wibowo and Lily D Li, 'The Effects of Over-Reliance on AI Dialogue Systems on Students' Cognitive Abilities: A Systematic Review' (2024) 11(1) *Smart Learning Environments* 28.

12 Chen and Sheehy (n 6).

13 Viktoriia Molodtsova et al, 'The Significance of Artificial Intelligence in Fostering Professional Competencies of the Future: A Systematic Review' (2025) 19(2) *Revista Eduweb* 280; Zhai, Wibowo and Li (n 11).

14 Nickolas James, 'The Truth(s) About AI and Legal Education: A Discourse Analysis of the Conflicting Narratives Regarding the Implications of Generative AI for the Teaching of Law' (paper presented at the 2023 Australasian Law Academics Association Conference, 2023); Karl de Fine Licht, 'Generative Artificial Intelligence in Higher Education: Why the "Banning Approach" to Student Use Is Sometimes Morally Justified' (2024) 37(3) *Philosophy & Technology* 113.

generation and opportunities for self-assessment.[15] Others urge greater caution, warning that the apparent convenience of GenAI may encourage cognitive offloading and weaken the development of rigorous, independent reasoning.[16] In legal education, that caution carries particular force. The surface fluency of GenAI output can obscure defects in authority, analysis and factual application, especially for novice students who have not yet developed the disciplinary schemas needed to identify those weaknesses.

For that reason, the central issue is not whether GenAI should simply be embraced or rejected. The more useful question is how legal education should respond to a tool that is likely to remain present in both legal study and legal practice, yet is unreliable in ways that matter professionally.[17] A strategy of simple denial may offer short-term reassurance, but it risks leaving students unprepared for the conditions in which they will later work.[18] Equally, unstructured adoption risks displacing the very effort through which legal competence is formed. In this respect, legal education faces a design problem rather than merely a policy problem.

Assessment is a particularly important site at which to address that design problem.[19] It is through assessment that students come to understand what matters in a subject, what standards govern performance and what kinds of intellectual work are rewarded. It is also the point at which concerns about academic integrity, learning, performance and professional formation converge most sharply. A focus on assessment therefore allows legal educators to move beyond abstract debates about whether GenAI is good or bad for education and instead ask a more precise question: can assessment be designed so that students learn to evaluate GenAI output against legal and professional standards?

That question is especially important in first-year law teaching. At that stage, students are still acquiring the conceptual and analytical foundations of legal reasoning, including the cognitive structures through which legal rules, concepts and factual patterns are organised and connected.[20] Early exposure to GenAI may therefore be educationally productive only if it is carefully scaffolded and explicitly critical. If introduced too early, or too permissively, GenAI may not simply save time; it may interrupt the cognitive effort through which foundational competence is built.[21] A staged approach is therefore preferable. Students should encounter GenAI in ways that preserve the

---

15 Xia et al (n 8).
16 Michael Veale et al (n 2) 1–3.
17 Koos and Wachsmann (n 11).
18 Xia et al (n 8).
19 Chen and Sheehy (n 6).
20 See Juan Diaz-Granados, 'Cognition, Language, and Law: Introspective Reflections on How we Learn, Teach, and Understand Law Across Languages and Legal Systems' (2023) 11(2) *International Journal of Law, Language & Discourse* 22.
21 Nataliya Kosmyna et al, 'Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task', *arVix* (Preprint, 2025) <https://arxiv.org/abs/2506.08872>.

primacy of independent legal reasoning while gradually developing the ability to use technological tools critically and responsibly.[22]

This article proceeds on that basis. It does not treat GenAI and traditional pedagogy as mutually exclusive,[23] nor does it assume that all forms of integration are equally desirable.[24] Instead, it examines one concrete assessment setting in which first-year law students were required to use GenAI, critique its output and evaluate its adequacy against disciplinary expectations. The significance of that design lies in its shift of emphasis. Rather than asking whether GenAI can produce acceptable legal work, it asks whether interaction with GenAI may provide students with opportunities to exercise evaluative judgment by comparing generated output against disciplinary and professional standards. That is the sense in which assessment matters here: not as a site of technological substitution, but as a site of disciplined comparison, critique and professional formation.

III AI AND LEGAL ASSESSMENT

Assessment is the most direct site at which legal education must respond to GenAI.[25] It shapes what students attend to, what they practise, and what they come to understand as competent performance. In the present context, assessment matters not only because GenAI creates new integrity concerns, but because it creates an opportunity to redesign tasks so that students engage directly with the strengths and weaknesses of machine-generated legal work. That is a materially different objective from asking whether GenAI can itself produce acceptable legal analysis. The pedagogical aim is not to replace human legal reasoning, but to use GenAI as an object of analysis through which professional judgment can be developed and exercised.

A central difficulty in this area is what Thomas Corbin, Phillip Dawson and Danny Liu describe as an 'enforcement illusion'.[26] Policies that merely tell students how they may or may not use GenAI can create the appearance of control without ensuring meaningful compliance or enforceability.[27] Institutions may announce a position and treat the problem as solved, even where monitoring is difficult, proof is uncertain and incentives for concealed use remain strong. In that setting, an assessment regime built primarily on prohibition risks

---

22 Repici (n 9).

23 For instance, the Socratic and Dialectic methods are not incompatible with the responsible use of AI: see Juan Diaz-Granados, Inteligencia Artificial y Educación Superior, 'Un Aliado o Un Enemigo?' [Artificial Intelligence and Higher Education: Friend or Foe?], *El Espectador* (27 November 2024) <https://www.elespectador.com/ciencia/opinion-inteligencia-artificial-y-educacion-superior-un-aliado-o-un-enemigo/>.

24 Chen and Sheehy (n 6).

25 Xia et al (n 8).

26 Thomas Corbin, Phillip Dawson and Danny Liu, 'Talk is Cheap: Why Structural Assessment Changes are Needed for a Time of GenAI' (2025) 50(7) *Assessment & Evaluation in Higher Education* 1087, 1095.

27 Ibid.

directing attention towards hidden use and detection rather than toward the quality of student judgment.

The present study was designed against that background. GenAI use was built into the task, and the rubric assessed students' critical engagement with the output they generated. This approach was therefore a structural rather than merely discursive response.[28] Instead of attempting to police invisible use, the assessment made GenAI visible and assessable. The educational significance of that move lies in the relocation of pedagogical attention: from rule enforcement to evaluative judgment, and from suspicion of hidden assistance to critical examination of what the tool actually produces.

This approach also clarifies what GenAI literacy should mean in legal education, especially in first-year subjects.[29] It should not be understood as technical expertise in model architecture or software design. Rather, at this level, GenAI literacy includes the capacity to formulate prompts with care, recognise unsupported or fabricated propositions, verify authorities independently, identify superficial analysis and understand that responsibility for legal advice remains with the human user. Properly conceived, GenAI literacy is therefore continuous with traditional legal skills rather than opposed to them. It extends the student's obligation to read carefully, reason accurately, verify authority and exercise judgment under conditions of uncertainty.

The broader higher education literature suggests that many teachers see some place for ChatGPT and related systems in teaching and learning, even while recognising substantial risks for assessment integrity and skill development.[30] One common response has been to shift back towards in-person or invigilated assessment formats.[31] Those strategies are understandable, but they do not by themselves resolve the deeper curriculum problem. A return to invigilated formats may manage certain integrity risks, yet it may also avoid the more difficult pedagogical task of teaching students how to use GenAI responsibly in the environments in which they are likely to encounter it beyond the classroom.

Although literature on GenAI-integrated assessment in legal education remains limited, adjacent scholarship in higher education is instructive.[32] David Ernesto Salinas-Navarro and others argue that GenAI can support authentic assessment when it is treated not merely

---

28 Ibid.

29 Ibid.

30 Chen and Sheehy (n 6); Xiaojing Weng et al, 'Assessment and Learning Outcomes for Generative AI in Higher Education: A Scoping Review on Current Research Status and Trends' (2024) 40(6) *Australasian Journal of Educational Technology* 37; see also Diaz-Granados (n 23).

31 Xia et al (n 8).

32 Weng et al (n 30). On the still-developing state of the legal-education literature specifically, see Steve Lorteau and Douglas Sarro, 'Artificial Intelligence in Legal Education: A Scoping Review' (2026) 60(1) *The Law Teacher* 138. On empirical and assessment-focused work in law, see Armin Alimardani, 'Generative Artificial Intelligence vs. Law Students: An Empirical Study on Criminal Law Exam Performance' (2024) 16(2) *Law, Innovation and Technology* 777; Stuart Hargreaves, '"Words Are Flowing Out Like Endless Rain into a Paper Cup": ChatGPT and Law School Assessments' (2023) 33 *Legal Education Review* 69.

as an 'agent-to-write' or 'agent-to-answer questions', but as an 'agent-to-think-about' intended learning outcomes and real-world scenarios.[33] That insight is especially valuable in law. Authentic legal tasks are not exhausted by the production of text. They require diagnosis of legal problems, selection and rejection of authorities, judgment about relevance and uncertainty, and communication tailored to a specific audience and professional purpose. An assessment that integrates GenAI is therefore pedagogically sound only if it engages students in those forms of judgment rather than concealing their absence behind superficially plausible output.

On that view, the incorporation of GenAI into legal assessment requires more than attaching a new tool to an old task. It requires assessment design that preserves the centrality of independent legal reasoning while using GenAI to illuminate what competent legal work involves. The present study adopts that position. It treats GenAI not as a surrogate legal writer, but as a comparator against which students may learn to identify weakness, defend disciplinary standards and better understand the distinctively human character of legal judgment.

IV METHOD

To address the research question, the study examined a two-part assessment conducted in a first-year Torts cohort of 120 law students. The activity was undertaken as a quality assurance evaluation of an assessment innovation, drawing on student feedback collected through a routine post-assessment debrief and on unsolicited written reflections.[34] The assessment design was informed by David Kolb's experiential learning cycle, which identifies four stages of learning: concrete experience, reflective observation, abstract conceptualisation and active experimentation.[35] Kolb's framework was selected because the task was not confined to the production of GenAI output. Rather, it required students to encounter that output, reflect on its adequacy, draw broader conclusions about its strengths and limitations, and then produce their own legal writing in light of those reflections. In that respect, the model provided an appropriate structure for investigating how students experienced GenAI within an assessment setting.

The assessment was structured around a Letter of Advice based on a provided fact pattern. Students were first required to use a GenAI tool to generate a draft of the advice and then critically evaluate that output. Drawing on this experience, students prepared their own polished Letter

---

[33] David Ernesto Salinas-Navarro et al, 'Using Generative Artificial Intelligence Tools to Explain and Enhance Experiential Learning for Authentic Assessment' (2024) 14(1) *Education Sciences* 1, 21.

[34] The project underpinning this study was assessed by the relevant institutional Human Research Ethics Committee and determined to be exempt from formal ethics review as a quality assurance activity. The Committee confirmed that the results may be published provided that all published material is de-identified; this article reports de-identified material accordingly. Institutional identifiers and the application reference number have been withheld to preserve author anonymity for peer review.

[35] David A. Kolb, Experiential Learning: Experience as the Source of Learning and Development (Pearson Education, 2nd ed, 2015) 4.

of Advice, demonstrating their understanding of the legal issues raised, their ability to identify and apply relevant legal principles, their capacity to critically assess the accuracy and quality of GenAI-generated legal analysis, and their ability to communicate their findings in a professional legal form. The assessment therefore integrated GenAI in a way that required students to interrogate and improve upon machine-generated work through their own legal reasoning, rather than using GenAI as a substitute for that reasoning.

Mapped against Kolb's cycle, the first encounter with the GenAI draft constituted the concrete experience. The requirement that students critique that draft corresponded to reflective observation, because it asked them to identify weaknesses, omissions, superficial reasoning and unsupported assertions in the AI-generated text. The broader conclusions students drew about the benefits and limits of GenAI in legal work reflected abstract conceptualisation. Finally, the production of their own Letter of Advice, informed by that comparison and critique, represented active experimentation. The design thus treated GenAI not as a substitute author, but as a pedagogical prompt capable of provoking evaluative judgment.

The assessment also had a practical educational purpose. It familiarised students with a technology that some had not previously used, while placing that familiarity within a controlled legal task that required verification, critique and independent responsibility. The exercise was intended to allow students to engage with GenAI without encouraging uncritical reliance on it. In this respect, the assessment design aligned with the article's broader claim that early encounters with GenAI should be supervised, structured and explicitly critical.

Data were collected from three sources after completion of the assessment sequence. First, approximately 60 students participated in in-class polling during a debrief session. Secondly, the debrief included short in-class discussion contributions from approximately 10 speakers. Thirdly, eight students sent unsolicited emails to the lecturer reflecting on their experience of the exercise. Of those eight emails, four were purposively selected for excerpting because they illustrated themes that also appeared across the broader data. The polling, discussion and written reflections were then considered together in order to identify recurring patterns in student responses to the use-and-critique assessment design.

The analysis was not intended as a full thematic analysis in the qualitative research sense, but as a pedagogically oriented synthesis of recurring student perspectives relevant to assessment design.[36] It is not intended to produce statistically generalisable findings or causal claims about the effects of GenAI on student performance.[37] Its more limited ambition is to identify themes in student experience that may inform assessment design and generate questions for further, more rigorous

[36] Andrea LaMarre and Kerry Chamberlain, 'Innovating qualitative research methods: Proposals and possibilities' (2022) 6 *Methods in Psychology* 100083.

[37] Ibid.

study.[38] That narrower methodological posture is appropriate to an emerging field in which legal educators are still working out how GenAI should be integrated into teaching and assessment. So understood, the study offers a bounded account of what occurred in one cohort and why those experiences may be relevant to the design of legal assessment more broadly.

V FINDINGS

After the assessment had been completed and grades returned, students were debriefed in a lecture attended by approximately 60 members of the cohort. The debrief began with in-class polling designed to elicit students' overall reactions to the requirement that they use GenAI as part of the assessment. The polling indicated a generally positive response to the exercise as an educational experience. Students appeared receptive to the inclusion of GenAI in assessment even where they remained cautious about the value of the tool itself. That distinction became clearer in the discussion that followed.

The polling was followed by open class discussion involving approximately 10 students who engaged in the discussion and provided oral feedback. Their comments suggested that while many students valued the exercise, their experience of the tool itself was mixed. Some described ChatGPT as useful for generating initial ideas or producing a preliminary structure. Others regarded it as of limited utility for the substantive legal task because, on closer inspection, it failed to address the core legal problem with sufficient precision. Several students also observed that the generated output did not reflect their own reasoning or voice. Where the tool was regarded as useful, its usefulness was more limited: it helped some students get started and reduced the anxiety associated with the blank page. In that sense, students tended to describe GenAI as more helpful in overcoming writer's block than in producing reliable legal analysis.

Following the in-class oral debrief, eight students sent unsolicited written reflections by email. These emails are treated here not as isolated anecdotes, but as part of the broader pattern of student feedback. Across the eight emails, three themes recurred: first, students experienced the exercise as valuable because it required critical engagement with GenAI output; secondly, they regarded the AI-generated draft as superficially plausible but substantively unreliable; and thirdly, many students ultimately valued the experience of writing the advice themselves because independent drafting felt both more

[38] Virginia Braun and Victoria Clarke explain that 'the "keyness" of a theme is not necessarily dependent on quantifiable measures – but rather on whether it captures something important in relation to the overall research question.' Virginia Braun and Victoria Clarke. 'Using Thematic Analysis in Psychology' (2006) 3(2) Qualitative *Research in Psychology* 77, 82.

trustworthy and more professionally meaningful. The quotations below are illustrative of those recurring themes.[39]

The first theme concerned the value of critical engagement as a learning exercise. One example of a recurring pattern, as one student wrote, was the following: 'I enjoyed it as a critical thinking exercise …. I did not use the output in my final draft which I wrote from scratch.' Another student wrote: 'I thought it was a very good idea and definitely gave me a better perspective of the WHYS of the assignment. (And why it is so important that lawyers are human).' These comments are notable because they identify what students perceived as educationally valuable about the task: not the quality of the AI-generated answer, but the discipline of evaluating it. From the students' perspective, the value of the exercise therefore lay in the process of comparison and critique. The value of the exercise, from this perspective, lay in sharpening the student's own judgment.

The second theme was that the AI-generated draft was often impressive at first glance but fragile under closer scrutiny. One student explained the general sentiment in the following terms:

> It was scary how it was able to generate [outputs]…. It was a matter of analysing it and being able to pull it apart and identify issues within it, which i found helpful in understanding mistakes I may have made in my own assessment. If it were to be included in other pieces of assessment, I think it would definitely be effective.

This response is significant because it captures both attraction and resistance. The student recognised the persuasive surface quality of the output, but also the value of dismantling that surface through legal analysis. The educational force of the exercise therefore appears to have lain in the contrast between fluency and adequacy.

The third theme was that, although GenAI could be helpful at the margins, students continued to regard the personal exercise of thinking, reasoning and writing as essential. Expressing a common view, one student wrote: 'I personally found the AI aspect of the assignment really helpful. … find it more fulfilling to write from scratch…. I felt much more comfortable with the answer I produced.' This mixed position was common across the data. Students did not necessarily reject the exercise or even the tool. Rather, they treated ChatGPT as something that might assist with beginnings, structure or confidence while still preferring independent drafting for the final work product.

Taken together, the polling, discussion and written reflections suggest a consistent pattern. Students generally valued the exercise of working with and critiquing GenAI, but they did not treat the AI-generated draft as a satisfactory substitute for their own legal reasoning and writing. The immediate pedagogical value of the intervention therefore appears to have lain less in AI-assisted production than in AI-assisted comparison and critique.

---

[39] The remaining five responses were reviewed as part of the synthesis but are not separately quoted because they substantially reflected the same recurring themes. No materially contradictory theme was identified in those responses.

VI DISCUSSION

The findings suggest that the inclusion of GenAI in the assessment was experienced by students as worthwhile, but not because they regarded the tool as a substitute for their own work. That distinction is central to the article's argument. A well-designed GenAI assessment may succeed not by encouraging reliance on the technology, but by helping students understand where its use may be helpful, where it is unreliable, and why any reliance upon it must remain disciplined and limited.

More specifically, the oral feedback indicated that GenAI could assist at the earliest stages of drafting, but that it struggled with the precision and nuance required for competent legal reasoning. Students repeatedly reported that the tool could generate a plausible starting point, yet often misrepresented or oversimplified important aspects of the task, particularly in applying legal principles to the facts. Many therefore discarded the AI-generated draft and wrote their Letters of Advice independently. That pattern is pedagogically significant for two reasons. First, it suggests that students recognised the continuing importance of developing their own legal writing and reasoning skills. Secondly, it indicates that critical exposure to GenAI may teach students not merely what the tool can do, but what competent legal work still requires from them. The educational value of the exercise appears, on this account, to have lain less in AI-assisted production than in the comparison between polished but inadequate output and the standards of professional legal analysis.

The polling data point in the same direction. Approximately half of those polled found the AI-generated draft somewhat useful, especially as a preliminary aid for organising ideas and identifying legal issues. The other half regarded the draft as of limited value, citing inaccuracies, hallucinations and superficial legal analysis. More strikingly, approximately three-quarters reported that they ultimately drafted their Letters of Advice without relying on the AI-generated text. That figure is important not because it shows that students rejected GenAI altogether, but because it suggests that structured critical exposure did not produce dependency. On the contrary, when students were required to compare AI output against their own developing judgment, most preferred to trust their own drafting and reasoning.

The written feedback adds further depth to that picture. Students did not merely say that the output was flawed; they identified the kinds of flaws that matter in legal writing: unsupported propositions, shallow application, failure to adapt analysis to the client's circumstances, and insufficient doctrinal care. Those are precisely the evaluative judgments that legal assessment seeks to cultivate. In that sense, the findings support a cautious but productive pedagogical model. GenAI appears to have some utility in legal education as a prompt and a foil. It may help some students begin. But its utility in producing doctrinally accurate, authority-based, audience-sensitive and analytically persuasive legal work appears limited.

Legal accuracy, thus, emerged as a particularly important dimension of students' evaluation of GenAI. Students identified inaccuracies, unsupported propositions and hallucinated or unreliable material as reasons not to accept apparently fluent output at face value. This reinforces the importance of independent source verification as a component of GenAI literacy in legal education. In legal assessment, accordingly, the ability to interrogate the accuracy and authority of generated content is itself an important object of learning.

This conclusion is broadly consistent with existing scholarship. Omar Tayan and others argue that students and instructors should recognise that GenAI systems do not themselves offer critical thinking and that students should therefore be encouraged to engage in critical analysis and problem-solving.[40] A similar point appears in legal education scholarship emphasising that students must first learn how to do legal work themselves before they can use GenAI effectively and responsibly.[41] That proposition resonates strongly with the responses in this study. The students who valued the exercise most did not describe themselves as liberated from legal work. They described themselves as better able to understand why they still needed to do it. The use-and-critique model therefore supports a developmental position: students may be introduced to GenAI, but only in ways that reinforce rather than replace independent competence.

The findings also align with broader work on assessment in higher education. Qi Xia and others argue that GenAI requires institutions to revisit program goals so as to cultivate self-regulated learning, responsible engagement with technology and academic integrity.[42] The present assessment responds to that challenge in one specific way. Rather than preserving older assessment forms unchanged, it builds interaction with GenAI into the task while making critique itself an assessable outcome. That is not a complete solution to the wider challenge GenAI poses for legal education, but it is an example of structural redesign rather than policy rhetoric alone.

A *Experiential Learning*

The findings may also be understood through the lens of Kolb's experiential learning cycle, which structured both the assessment design and the analysis. The concrete experience stage was realised when students engaged directly with GenAI to produce a draft Letter of Advice. Student reactions to the tool's apparent fluency capture the force of that first encounter. Reflective observation then occurred as students compared the generated draft with their own developing understanding of the task and began to identify defects in authority, reasoning and factual application. This reflective stage was

---

40 Omar Tayan et al, 'Considerations for Adapting Higher Education Technology Courses for AI Large Language Models: A Critical Review of the Impact of ChatGPT' (2024) 15 *Machine Learning with Applications* 1, 12.

41 Repici (n 9) 2.

42 Xia et al (n 8).

pedagogically important because it required students not merely to state that the answer was weak, but to articulate how and why it was weak.

Abstract conceptualisation is visible in students' ability to derive broader conclusions from that experience. Their comments did not remain confined to one flawed output from one tool. Rather, students drew more general lessons about the relationship between language generation and professional judgment, including the continuing importance of human legal reasoning. Active experimentation, finally, was manifested in students' subsequent drafting choices. Having reflected on the limits of the AI-generated draft, many wrote their own advice from scratch or used the generated text only minimally. On this reading, the task did more than expose students to a technology. It channelled that exposure into revised action and hence into experiential learning about GenAI's capabilities and constraints.

B *Cognitive Load and Novice Learners*

Cognitive load theory offers a complementary, though more tentative, explanation for the divergent student responses.[43] The task itself involved substantial intrinsic load:[44] students were required to understand tort doctrine, apply legal reasoning to a fact pattern, and communicate in a professional genre. For some students, particularly those with less developed legal schemas, the AI-generated draft may have reduced initial difficulty by offering structure and helping them begin. That possibility is consistent with the comments describing GenAI as helpful for getting started or overcoming the blank page.

For other students, however, the tool appears to have imposed extraneous load.[45] Instead of reducing difficulty, it created additional work by requiring them to identify and correct inaccuracies, hallucinations and superficial analysis. The description of GenAI as a 'time-wasting trap' reflects that experience well. Importantly, however, the assessment was designed not merely to expose students to the tool, but to require evaluation of it. In that sense, the critical reflection component may be understood as generating germane load:[46] cognitive effort directed toward schema development and deeper understanding. Student comments suggesting that analysing the AI draft helped them see possible weaknesses in their own work support that interpretation. This is best understood as an interpretive explanation rather than a measured finding, but it helps explain why the same tool was

---

43 See John Sweller, 'Cognitive Load Theory' in Jose P Mestre and Brian H Ross (eds), *The Psychology of Learning and Motivation: Cognition in Education* (Academic Press, 2011). For empirical evidence of GenAI's differential effect by student skill level — substantial gains for lower-performing students and performance declines for higher-performing students — see Jonathan H Choi and Daniel Schwarcz, 'AI Assistance in Legal Analysis: An Empirical Study' (2025) 73 *Journal of Legal Education* 384.

44 John Sweller distinguishes intrinsic load (inherent task complexity), extraneous load (unnecessary cognitive burden from poor instructional design), and germane load (productive effort directed at schema construction): Sweller (n 43) 37.

45 See above n 44.

46 See above n 44.

experienced as supportive by some students and counterproductive by others.

C *Evaluative Judgment and Professional Formation*

As the findings are directly relevant to evaluative judgment, they also have implications for professional formation. Legal advice is not simply a collection of legally themed sentences. It is a professional act directed to a client, grounded in authority, and shaped by judgment about relevance, uncertainty, risk and communicative purpose. Several students appeared to recognise that more sharply because the exercise required them to compare human and machine-generated drafts. The inadequacy of the AI output became, in effect, a way of seeing what legal professionalism consists of.

That point connects closely with scholarship on evaluative judgment. Joanna Tai and others define evaluative judgment as the ability to judge the quality of one's own work and the work of others, and argue that it should be treated as a central goal of higher education.[47] The present assessment can be understood as cultivating that capacity through an unusual comparator: not another student's draft, but a machine-generated text with persuasive surface fluency and hidden weaknesses. When students identified omissions, weak authority use and poor factual application in the AI-generated output, they were practising evaluation in a way closely tied to professional legal standards. On that view, the pedagogical significance of the exercise extends beyond AI literacy narrowly understood. By requiring students to assess generated legal work against disciplinary standards, the task provided an opportunity to practise a form of evaluative judgment that is also relevant to the development of professional legal judgment.

VII LIMITATIONS AND IMPLICATIONS

The study has several limitations. First, it is a single-site study based on one first-year Torts cohort. Secondly, it relies substantially on self-reporting. Students may have misunderstood the task, misdescribed their experience or selectively recalled what they found useful or unhelpful. Thirdly, the polling and discussion took place in a lecture attended by only part of the cohort. Approximately 60 of 120 students participated in the debrief, which means that the oral data reflect a self-selected subset of the class. Students who attend debriefs and volunteer views may differ systematically from those who do not. These design constraints suggest future research directions for innovative research.

A further limitation is the absence of demographic or background analysis. The present study does not disaggregate responses by age, prior education, language background, prior GenAI use, digital confidence or academic performance. Those factors may plausibly

---

[47] Joanna Tai et al, 'Developing Evaluative Judgement: Enabling Students to Make Decisions about the Quality of Work' (2018) 76(3) *Higher Education* 467.

shape student experience and would provide areas for fruitful investigation in future studies.

Finally, as the study is exploratory in nature, it neither aims to support nor provides a basis for causal inference, institutional generalisation, or broad claims about all GenAI tools or all legal education settings. Its value lies instead in offering a carefully bounded account of one pedagogical intervention and the student responses it elicited. So framed, the implications are modest but useful. The findings suggest that structured use-and-critique assessment may provide one productive way of introducing GenAI into legal education, particularly where the aim is to cultivate evaluative judgment rather than automate student performance. They also suggest that future research should examine more systematically how such designs operate across different cohorts, stages of study and assessment formats.

VIII CONCLUSION

This article has argued that a promising response to GenAI in legal education is neither blanket prohibition nor uncritical adoption, but a structured use-and-critique model embedded in assessment. The exploratory findings reported here suggest that such a model can expose students to GenAI while also requiring them to evaluate its adequacy against legal and professional standards. In that respect, the value of the intervention lay less in AI-assisted drafting than in the development of evaluative judgment through comparison, critique and independent legal writing.

The student responses revealed a consistent pattern. Many students recognised that GenAI could be useful in limited ways, particularly in generating an initial structure or helping them begin. At the same time, they remained sceptical of its capacity to produce reliable legal analysis, especially where the task required doctrinal accuracy, careful factual application and audience-sensitive advice. That dual response is central to the article's contribution. Students did not need to be shielded from GenAI. They needed to be taught how to use it critically, and how to recognise the difference between persuasive surface fluency and competent legal work.

The study therefore provides exploratory evidence consistent with a staged and supervised approach to GenAI in legal education, especially in early-year subjects. On this view, GenAI may have a supplementary role in supporting learning, but it should not displace foundational legal skills such as analysis, source verification, legal reasoning and professional writing.[48] The use-and-critique model is one way of preserving that balance. It does not ask students merely to produce legal text with technological assistance. It asks them to explain why one form of legal writing is better than another, and in doing so strengthens capacities that are central both to legal education and to legal practice.

[48] As one of the authors explains, 'law should orient technology toward human—not machine—flourishing.' Diaz-Granados (n 1) ix.

The article's broader implication is modest but important. If GenAI is likely to remain part of the professional environment into which law students graduate, then legal education must do more than attempt to exclude it. It must develop assessment designs that make visible the limits of GenAI as well as its attractions, and that cultivate disciplined, critical and professionally responsible engagement with it. The present study does not offer a comprehensive solution, nor does it provide a basis for broad generalisation. It does, however, suggest that assessment can be redesigned so that interaction with GenAI provides students with opportunities to practise evaluating generated legal work against disciplinary and professional standards, rather than simply using the technology as a shortcut to produce legal work.

More specifically, the exercise appears to have prompted students to reflect on a core feature of legal work: its dependence on judgment. Legal advice is not merely text generation.[49] It involves authority, interpretation, uncertainty, relevance and responsibility to clients. By comparing their own work with machine-generated drafts, students were placed in a position to see those features more sharply. In that sense, the exercise contributed not only to GenAI literacy, but also to a clearer appreciation of the human dimensions of legal professionalism.[50]

Ultimately, the significance of the study lies in what it suggests about assessment design. GenAI need not be treated only as an integrity problem or as a substitute for student performance. It can instead be used as a comparator against which students learn to identify weakness, defend disciplinary standards and practise evaluative judgment. For legal educators, that may be the more productive path: not asking whether GenAI can do legal work for students but designing assessment so that students learn more clearly what good legal work requires.

---

[49] Joshua Krook et al, 'Large Language Models (LLMs) for Legal Advice: A Scoping Review' (Research Paper, SSRN, 2024) <https://ssrn.com/abstract=4976189>; Q Li and Q Zhang, 'Court Opinion Generation from Case Fact Description with Legal Basis'(2021) 35(17) *Proceedings of the AAAI Conference on Artificial Intelligence* 14840; Xia et al (n 8).

[50] Julian Webb argues that '[t]he idea of the democratic intellect implies an education of humane breadth, grounded in the liberal arts and with a strong philosophical footing. … it means that students are indeed enabled to learn about the values and principles that underpin the discipline of law – a step that, in itself, requires us to take values much more seriously than we do now.' Julian Webb, 'Taking Values Seriously: The Democratic Intellect and the Place of Values in the Law School Curriculum' in Michael Robertson, Lillian Corbin, Kieran Tranter and Francesca Bartlett (eds), *The Ethics Project in Legal Education* (Routledge, 2010) 24.