ASIA PACIFIC JOURNAL
OF EDUCATORS AND EDUCATION

Volume 41, Issue 1, 2026
https://doi.org/10.21315/apjee2026.41.1.6

Research Article:

Evaluating the Role of ChatGPT as a Legal Educational Companion

Jenita Kanapathy[1]* and Jasreena Avelyn Shamini[2]

[1]School of Law and Governance, Faculty of Business and Law, Taylor's University, No. 1, Jalan Taylor's, 47500 Subang Jaya, Selangor, Malaysia
[2]Faculty of Language and Communication, Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim, Perak, Malaysia

*Corresponding author: Jenita.Kanapathy@taylors.edu.my

ABSTRACT

This study evaluates the efficacy of ChatGPT as a study companion in legal education, addressing ongoing debates regarding its pedagogical benefits and potential risks to critical thinking. Guided by the IDEE framework, a mixed-methods approach (quantitative and thematic qualitative analysis) was adopted using a pre-test/post-test design involving 64 first-year law students. Student performance was measured through a 10-question multiple-choice quiz, administered without ChatGPT in the pre-test and with ChatGPT in the post-test, with supplementary qualitative feedback collected via an open-ended questionnaire. The findings reveal a statistically significant improvement in overall student performance, particularly in facilitating progression from failing to passing and good grades. However, no significant evidence was found to suggest that ChatGPT supports the attainment of the highest academic grades (A or A–). This suggests that ChatGPT is effective in supporting lower-order cognitive processes but has limited impact on higher-order reasoning required for advanced legal analysis. These findings should be interpreted with caution due to methodological limitations, including the absence of a control group and potential test familiarity effects. Overall, ChatGPT functions best as a supplementary learning tool that supports foundational understanding, while human instruction remains essential for developing critical thinking and nuanced legal judgement.

Keywords: ChatGPT, Generative AI, mixed methods, legal education, student performance

Published: 9 June 2026

To cite this article: Jenita, K., & Jasreena, A. S. (2026). Evaluating the role of ChatGPT as a legal educational companion. *Asia Pacific Journal of Educators and Education, 41*(1), 111–131. https://doi.org/10.21315/apjee2026.41.1.6

INTRODUCTION

The growth of artificial intelligence (AI) is drastically changing job opportunities where education serves, raising concerns about what and how to teach future generations. Such issues underscore the importance of education to provide the future workforce with the necessary skills and competencies to survive in a rapidly changing world (Zhai, 2021). Despite the rapid proliferation of generative AI tools in education, existing studies have largely focused on general higher education contexts such as STEM and language learning, with limited attention to discipline-specific applications that require specialised reasoning, such as legal education (Kasneci et al., 2024; Dwivedi et al., 2023). Therefore, it is important to understand the capabilities of AI tools designed to replace human efforts to save both money and time. Natural Language Processing (NLP) is an example of AI that has transformed many aspects of the modern world. Researchers have developed conversational AI systems like ChatGPT (OpenAI, n.d.) to provide conversational responses to human queries, especially given the availability of huge databases (Lock, 2022; OECD, 2021). Developed by OpenAI, it is an advanced AI language model designed to generate human-like text in response to input. It employs a deep learning technique called a transformer and draws from diverse text sources for coherent responses. This versatile tool finds application in natural language processing tasks, including answering questions, generating text, explaining concepts, and interactive conversations. Students can engage in learning activities, personalise their learning experience, streamline the enrolment process, and increase tutor/lecturer efficiency (Okonkwo & Ade-Ibijola, 2021). These studies collectively suggest that generative AI may enhance accessibility, engagement, and personalised learning. However, they primarily emphasise functional benefits rather than examining deeper pedagogical implications, such as the development of higher-order cognitive skills. According to Yin et al. (2021), students' motivation and performance improve when they use chatbot-based microlearning systems. Additionally, AIs such as ChatGPT are noted for developing critical reasoning and writing skills in students (Irfan et al., 2023). From a pedagogical perspective, these claims may be situated within constructivist learning theory, which emphasises active knowledge construction through interaction and feedback. Generative AI tools such as ChatGPT can potentially function as scaffolding mechanisms that support learners in constructing understanding (Frank, 2024; Hwang et al., 2020; Tran et al., 2025).

However, while the argument thus far points toward AI's good effect, the reality is mixed, especially in education. Emerging scholarship highlights a tension between efficiency gains and potential risks, including over-reliance, reduced cognitive effort, and challenges to academic integrity (Cotton et al., 2024; Kasneci et al., 2023). As chatbots have become popular in recent years, the debate about their impact on education and their use as student engagement tools is gaining a lot of attention (Miller, n.d.). Alarie (2023) argued that AIs such as ChatGPT cannot replace critical thinking. The arguments brought in by Stott and Stott (2023) are of relevance. They provided their perspective on the use of AI in tax education. They noted that AI fails to match humans in making appropriate judgements and reasoning capabilities. They particularly argued that AI is not able to provide context-

specific feedback. The second point is that AIs such as ChatGPT's ability to respond to queries especially regarding tax-related topics are limited. It could be that AI lacks understanding of tax language or jargon which is required before providing customised responses. Lastly, the authors pointed out that AIs such as ChatGPT are unable to discern the sensitivity of a piece of information or data. Hence this argument can be applied to legal education, where legal language and jargon are in a similar context with tax education.

Despite several studies on the use of AI in the education of diverse disciplines, the extant literature is not rife with such studies in legal education. This is particularly significant given that legal education is inherently distinct from other disciplines, as it requires not only knowledge acquisition but also the development of analytical reasoning, argumentation, and interpretation of complex legal texts (Sullivan et al., 2007). Existing studies have not sufficiently examined whether generative AI tools can support these higher-order cognitive processes or merely facilitate surface-level learning. Furthermore, there is limited empirical research employing mixed method approaches to evaluate both performance outcomes and student perceptions within legal education contexts. Hence, this study investigates the efficacy of ChatGPT in legal education.

LITERATURE

The Capabilities and Potential Benefits of ChatGPT in Education

The integration of Generative AI (GenAI) into educational domains most notably through Large Language Models (LLMs) like ChatGPT promises a paradigm shift in academic productivity as the current discourse remains heavily weighted toward utilitarian efficiency. Zhai (2022) demonstrates this by highlighting the tool's capacity to produce coherent, systematic academic papers in a fraction of the time required by human researchers. However, this efficiency raises a fundamental epistemological concern. If the researcher's role is reduced to minor editing and reorganisation (Zhai, 2022), the traditional process of knowledge construction which relies on the iterative, cognitive struggle of drafting and refining ideas is bypassed. This suggests a potential shift from constructionist learning to a curatorial model, where the student or researcher manages outputs rather than deeply engaging with the subject matter.

In specialised fields such as medical education, the role of ChatGPT as a virtual co-teacher (Lee, 2024) further emphasises accessibility and real-time feedback. By streamlining administrative tasks and augmenting visual programs like histology slides, LLMs ostensibly free students to focus on practical skills (Lee, 2024; Kasneci et al., 2023). These benefits are further reinforced by survey data from Trakia University, where professors view the technology as a vital tool for stimulating creativity and managing time-intensive tasks (Kiryakova & Angelova, 2023). From a pedagogical standpoint, these benefits may be interpreted through constructivist learning theory, where knowledge is actively constructed through interaction and feedback. ChatGPT may function as a form of scaffolding,

supporting learners within their zone of proximal development (Frank, 2024; Hwang et al., 2020; Tran et al., 2025). However, whether such scaffolding promotes independent learning or fosters over-reliance remains contested. Chan (2023), on the other hand, proposes an AI education policy for higher education to address the multifaceted implications of AI integration in university teaching and learning. According to a study at the University of Pennsylvania, ChatGPT fared well on a business management exam, answering basic operations management and process-analysis questions at the Wharton Business School where it earned a B to B– grade (Terwiesch, 2023). By integrating ChatGPT, instructors can streamline tasks, saving valuable time, while students experience a more tailored and adaptive learning experience (Javaid et al., 2023).

Despite these functional advantages, the literature largely adopts a utilitarian perspective that prioritises output over cognitive process. The transition from a B– grade performance (Terwiesch, 2023) to genuine expertise requires more than just efficient scaffolding. It requires higher-order cognitive skills such as analysis, evaluation, and synthesis (Kasneci et al., 2023; Cotton et al., 2024). If learners blindly trust generated content (Kiryakova & Angelova, 2023), the risk is an epistemological erosion where the ability to verify authenticity forming a cornerstone of scientific inquiry is sacrificed for speed. Hence, there remains a critical need to examine how these tools influence the construction of new knowledge. Without a framework to safeguard evaluative rigor, the future of AI in education risks producing a generation of learners who are efficient consumers of information, but lack the deep analytical skills required to contribute original insights to their fields.

AI in Legal Education and Professional Practice

ChatGPT was also explored for its potential use in legal education and legal practice. Legal education differs fundamentally from other disciplines due to its emphasis on doctrinal reasoning, interpretation of complex texts, and argumentative analysis. As such, the integration of generative AI raises unique pedagogical questions regarding whether such tools can genuinely support legal reasoning or merely assist in surface-level content generation (Sullivan et al., 2007; Dwivedi et al., 2023). Perlman (2023) demonstrated the implications effectiveness of ChatGPT and its more advanced version, Bing GPT in a study about the implications of using AI in legal services and society. This study was published on the website of the Harvard Law School, Centre on the Legal Profession (Perlman, 2023). In this study, Perlman also relied on ChatGPT to write a paper, except for the abstract, the preface, the outline headers, the epilogue, and the prompts used to garner the responses from ChatGPT or Bing GPT. Perlman (2023) also conducted tests on Bing GPT, which was an advanced version of ChatGPT that was released in February 2023 for beta testing. In the said test, Perlman (2023) asked Bing GPT, 15 challenging legal analysis questions about legal ethics that were formatted as multiple-choice questions. Bing GPT was able to answer 12 out of the 15 legal analysis questions correctly and was assessed as performing at a B+/B law student level (Perlman, 2023) as well as anticipated to improve with time. According to Ajevski et al. (2023), keeping up with technological advancements is no longer an option for law schools, and ChatGPT provides a fresh impetus to review the

curriculum to ensure that it supports students in transitioning into new working practices. Ajevski et al. (2023) contend that law schools should create measures to discourage students from using generative AI to pass assessments while also teaching them how to utilise it productively in the profession. A study conducted at the University of Minnesota Law School found that ChatGPT passed all four classes based on its final exam, averaging a C+ across all exams, and that ChatGPT's grades could be enough to graduate from law school (Choi et al., 2022). Study findings revealed that ChatGPT performed best in exclusively essay-based assessments, with questions that encouraged students to reflect upon or analyse relatively broad topics (Hargreaves, 2023). There is some guidance from scholars on how to approach ChatGPT problems in legal exams (Weiler, 2023). Perlman (2023) views that first-year law students can benefit by learning about using ChatGPT as a tool in their clinical programs as well as for legal research and writing. A study to test the ability of ChatGPT to assist with common tasks performed by law professors showed that ChatGPT can provide law professors with near-finished products for routine tasks and a solid jumping-off point for those that are more complex (Oltz, 2023). Taylor (2023), on the other hand, highlights the pros and cons of ChatGPT for Law Firms based on the view provided by ChatGPT. Ward (2023) reveals that while ChatGPT can assist in organising concepts and improving the clarity of explanations, some educators are concerned that it may hinder students' development of strong writing skills, meanwhile, despite its limitations, legal professionals believe that tools like ChatGPT can enhance contract drafting and improve efficiency in the legal industry. A study to test ChatGPT's abilities in the legal academic realm, assigning exercises and quizzes designed for law and computer science graduate students by an adjunct law professor in Texas, remarkably showed that ChatGPT not only provided competent answers relevant to the legal topic but also demonstrated interdisciplinary knowledge by mentioning its application in another field (Ball, 2023). Collectively, these studies suggest that while ChatGPT demonstrates competence comparable to average law students in structured assessments, its ability to replicate nuanced legal reasoning and contextual judgement remains limited. This indicates a potential gap between performance outcomes and deeper cognitive engagement in legal learning.

Concerns and Challenges: Academic Integrity and Critical Thinking

However, the literature review reveals that some downside concerns regarding ChatGPT countervail its potential benefits. These concerns reflect broader tensions identified in the literature between efficiency and epistemic integrity, where the convenience of AI-assisted learning may undermine authentic knowledge construction and independent cognitive effort (Cotton et al., 2024; Kasneci et al., 2024). The first of these is the abuse of ChatGPT which leads to rampant cheating and plagiarism. According to Stokel-Walker, in the event a student generates a research paper using ChatGPT, this would be tantamount to a violation of academic integrity (Stokel-Walker, 2023). In research conducted by Ventayen (2023) at a university in the Philippines, it was reported that nearly 83% of the students interviewed were aware of ChatGPT and most of them knew that ChatGPT-generated contents are not easily detectable by anti-plagiarism software (Ventayen, 2023). In addition, the majority of the students were found to be tempted

to cheat as a result of stiff competition to get excellent grades and land a better-paying job. However, when queried if they used ChatGPT to complete their assignments, many denied using ChatGPT in their assignments. However, the researcher suspects that the answers they provided were not honest. Secondly, there are concerns about the erosion of critical thinking in students as shared by an article on the PennCareyLaw website of the University of Pennsylvania (Riermaier, 2023). Zhai (2021) also observed in his research that ChatGPT is capable of handling essay writing and constructed response tasks easily which are normally used to evaluate students and students can easily outsource these tasks to ChatGPT. These general writing skills are easily conducted as compared to creativity and critical thinking using ChatGPT. Therefore, Zhai (2021) suggests that teachers should utilise alternative assessment formats which include creativity and critical thinking. This notion is also supported by Lee concerning exploring the potential use of ChatGPT in medical education. Lee (2024) believes that a more rigorous examination of the influence of ChatGPT and other AI technologies on student learning outcomes such as knowledge retention, problem-solving abilities, and critical thinking skills is necessary. Lo (2023) also emphasised that assessment methods and institutional policies in schools and universities be updated immediately. Klingensmith (2023) views that although ChatGPT is capable of performing well enough to be comparable to the average law school student or an average law school professor, it is not able to provide the level of service that an average lawyer would provide. This literature suggests that ChatGPT is revolutionary and a game-changer as a virtual co-teacher to assist students in essay writing skills, but its efficacy in terms of developing creativity in problem-solving and critical thinking in students remains unclear. From the perspective of Bloom's Taxonomy, ChatGPT appears to effectively support lower-order cognitive processes such as remembering and understanding but may be less effective in fostering higher-order skills such as evaluation and creation, which are central to legal education.

Theoretical Framework: The IDEE Model

Given the complex and conflicting perspectives on integrating generative AI, a structured approach is necessary, such as the IDEE Framework. This study adopts the IDEE framework to structure its inquiry into ChatGPT's efficacy as suggested by Su and Yang (2023). The utility of the IDEE framework as a systematic model for integrating AI has been further demonstrated in diverse fields, including its use for incorporating ChatGPT into online discussions in literacy courses (Hu & Lin, 2025) and for implementing and evaluating Large Language Models in chemical engineering education (Keith et al., 2025). The selection of this framework helps ensure that the evaluation of ChatGPT is grounded in a systematic and pedagogically focused model, moving beyond a purely technical assessment. While the IDEE framework provides a structured model for implementation, it has not been extensively applied in empirical studies in legal education contexts. This study, therefore, extends the application of the IDEE framework by integrating it with empirical performance data and student perceptions. The IDEE framework provides a four-step guide for using Generative AI in education, ensuring a systematic and ethical implementation, which is crucial for a new technology like ChatGPT in a sensitive field like law.

The IDEE framework entails a 4-step process which is firstly to identify the desired outcome, secondly to determine the appropriate level of automation, followed by ensuring ethical consideration, and finally, to evaluate the effectiveness of ChatGPT. In line with the IDEE framework, the identified desired outcome of this study is to investigate whether ChatGPT would potentially improve student's performance in a multiple-choice quiz. Secondly, this study only employed a partial level of automation as students need to manually feed appropriate or relevant textual inquiry into ChatGPT based on test questions to garner a suitable reply from ChatGPT. To ensure ethical consideration as espoused by the IDEE framework, the quiz questions were administered in a random sequence for individual students using Taylor's Integrated e-Learning System (myTIMeS). Therefore, mitigating any cheating by students during the quizzes. Finally, statistical methods were used to evaluate the effectiveness of using ChatGPT as a study companion in legal education as suggested by the IDEE framework. A visual representation of the IDEE framework is depicted in Figure 1 illustrates the application of the IDEE framework in this study, showing the steps from identifying outcomes to evaluating effectiveness.

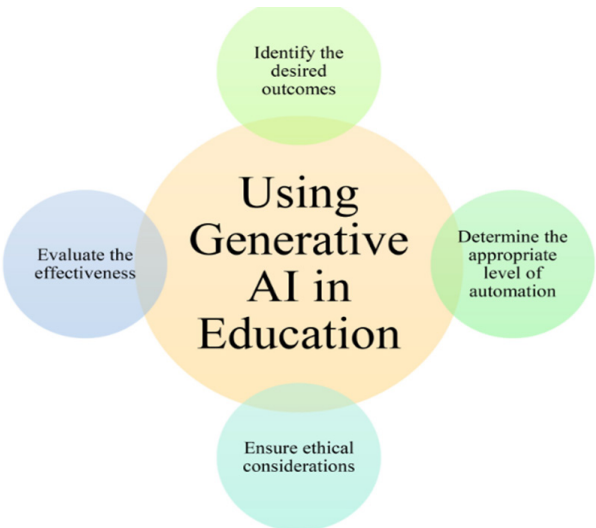

Figure 1. IDEE theoretical framework for using Generative AI in education (Su & Yang, 2023)

Research Gap and Study Motivation

The existing studies tend to focus either on performance outcomes or on conceptual discussions, with limited integration of quantitative and qualitative perspectives. Furthermore, there is insufficient evidence examining whether improvements in performance translate into meaningful learning in discipline-specific contexts such as law. Hence, this study addresses the following research questions:

RQ1: Does the use of ChatGPT improve the overall performance of legal students' marks?

RQ2: Does the use of ChatGPT improve performance from passing to good grades for legal students?

RQ3: Does the use of ChatGPT ensure the achievement of high grades (grade A or A–) for legal students?

RQ4: Does ChatGPT potentially serve as a companion for legal students?

METHOD

The study adopted a mixed-methods approach, employing a quantitative pre-test and post-test research design to assess student performance in a multiple-choice quiz (RQ1, RQ2, and RQ3). The use of a mixed-methods design enables both the measurement of performance outcomes and the exploration of students' learning experiences, thereby providing a more comprehensive understanding of the pedagogical impact of ChatGPT (Creswell & Plano Clark, 2018). The pre-test/post-test design is commonly used in educational research to evaluate intervention effects by comparing performance before and after exposure to a learning tool (Dimitrov & Rumrill, 2003). The quiz consisted of 10 questions, divided into 5 direct and 5 problem-based questions. Administered through myTIMeS, the quiz aimed to evaluate students' understanding of Promissory Estoppel. Taylor's University is one of the world's top 1% universities and the top 1 private university in Southeast Asia for 5 years in a row (2025 QS World University Rankings). In addition, it ranks 27 in Asia (2026 QS Asia University Rankings) ensuring a high-quality learning environment. The myTIMeS automatically marked and graded the multiple-choice questions, allowing students to view their results post-submission. The study involved 64 first-year LLB law students participating in the Contract Law module. The use of a single cohort reflects a convenience sampling approach, which, while appropriate for exploratory classroom-based research, may limit the generalisability of the findings to broader legal education contexts (Etikan et al., 2016). It should be emphasised that the students had previously received a lecture on the quiz topic, Promissory Estoppel, before the quiz. Therefore, students were expected to be familiar with the topic Promissory Estoppel to solve the multiple-choice quiz and were informed that quiz results would not impact their overall module grade, preventing coordinated efforts for the ChatGPT post-test. Ethical considerations were addressed by ensuring that participation did not affect students' academic grades and that responses were anonymised to protect student confidentiality.

For the pre-test, students were only permitted to refer to their lecture notes, but for the post-test, they were allowed to utilise both lecture notes and ChatGPT. The same quiz questions for the pre and post-tests were administered using myTIMeS. However, the use of identical questions may introduce a testing or familiarity effect, whereby improvements in post-test performance may be partially attributed to prior exposure rather than the

intervention itself (Campbell & Stanley, 1963; Roediger & Butler, 2011; van Gog & Sweller, 2015; Yang et al., 2021). The myTIMeS was set up to automatically deliver quiz questions in an arbitrary order to each student to reduce copying during the quiz, and the solutions were not revealed to the students. Both tests were administered back-to-back in one session under identical conditions (time limit, environment), with the only difference being the availability of ChatGPT in the post-test as an educational companion tool. While efforts were made to standardise testing conditions, internal validity may still be influenced by factors such as short-term memory retention and test-retest effects, which are inherent limitations of quasi-experimental designs (Shadish et al., 2002). To gather supplementary qualitative insights on the experience of using the tool (RQ4: ChatGPT as a companion), an open-ended question was administered to all students immediately following the post-test. The qualitative responses were analysed using a thematic analysis approach, allowing patterns and recurring themes in students' perceptions to be systematically identified and categorised (Braun & Clarke, 2006). Students were asked to provide feedback on their experience with ChatGPT, including its helpfulness and any challenges encountered during the quiz. This qualitative element was captured via a post-quiz questionnaire integrated into myTIMeS and was used to provide a nuanced interpretation of the quantitative results in the discussion section. However, the absence of a control group in this pre-test/post-test design limits the strength of causal claims, as observed gains cannot be attributed solely to the use of ChatGPT. This limits causal inference, as improvements in performance may also be influenced by confounding variables such as prior knowledge, test familiarity, or short-term learning effects rather than the independent use of ChatGPT. This exploratory study design was chosen to primarily assess the impact of the tool in a realistic learning scenario.

RESULTS

Overall Test Scores: Sample Characteristics

The SPSS software was used to analyse the quantitative data, and the paired sample *t* statistical test was employed for the quantitative analysis of the overall test scores. However, for the paired sample *t*-test to be valid, the difference of the paired values should be approximately normally distributed. For the overall test scores, the result of a Shapiro-Wilk test ($p = 0.076$) demonstrated that the samples were approximately normally distributed. In addition to the Shapiro-Wilk's test, the visual inspection of the histogram of the difference values between the pre-test and post-test scores as given by the SPSS software also showed that the samples were normally distributed with skewness of –0.177 (SE = .299) and kurtosis of 0.161 (SE = .59) as shown in Figure 2.

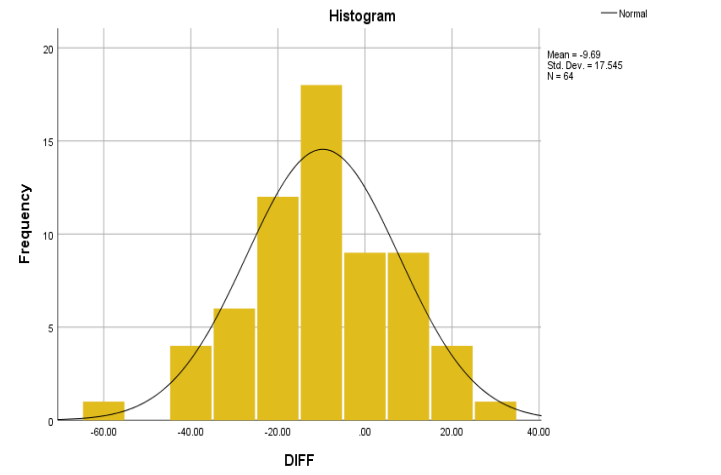

Figure 2. Histogram of the difference values between the pre-test and the post-test scores

The normal Q-Q plots as shown in Figure 3 also confirm that the difference of the pre-test and post-test scores are approximately normally distributed.

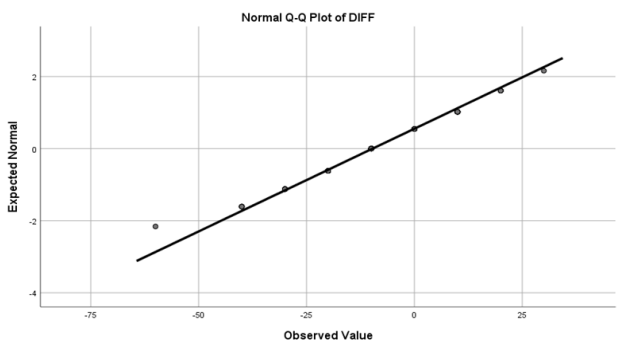

Figure 3. Normal Q-Q plot of the difference values between the pre-test and post-test scores

Tables 1 and 2 present the frequency counts and frequency percentages for the students' pre-test and post-test scores. The data from the frequency counts demonstrates that by using ChatGPT the quiz scores of students have significantly improved. It was observed that the percentage of students who attained 40 marks and below in the pre-test was 70.3% had reduced to 48.4% in the post-test. This suggests that ChatGPT plays a pivotal role as an educational companion, enhancing students' learning in legal education.

Table 1. Frequency table for pre-test scores

Scores	Frequency	Frequency (%)	Cumulative (%)
0	1	1.6	1.6
10	2	3.1	4.7
20	12	18.8	23.4
30	15	23.4	46.9
40	15	23.4	70.3
50	12	18.8	89.1
60	5	7.8	96.9
70	2	3.1	100.0
Total	64	100.0	

Table 2. Frequency table for post-test scores

Scores	Frequency	Frequency (%)	Cumulative (%)
10	1	1.6	1.6
20	1	1.6	3.1
30	11	17.2	20.3
40	18	28.1	48.4
50	15	23.4	71.9
60	12	18.8	90.6
70	6	9.4	100.0
Total	64	100.0	

The paired sample *t*-test used to assess the effect of ChatGPT on students' overall quiz scores revealed a statistically significant improvement from pre-test (M = 36.72, SD = 14.913) to post-test (M = 46.41, SD = 13.495), $t(63) = -4.417$, $p = .000$ (2-tailed). The mean increase in test scores was 9.687, with a 95% CI ranging from –14.070 to –5.305. The effect size calculated using Cohen's *d* was 0.68, indicating that ChatGPT had a medium effect on student test results.

Direct Questions (DQ) Quiz Scores: Sample Characteristics and Results

As highlighted earlier, there were five direct multiple-choice quiz questions in both the pre-test and post-test. The normality test indicated that the samples were not normally distributed. Hence, the Wilcoxon Signed Rank Test was used as a nonparametric alternative to the paired-samples *t*-test to evaluate the influence of ChatGPT on the five direct quiz questions. To assess if there were any changes to the quiz scores when students used ChatGPT as an additional learning tool to answer the direct multiple-choice quiz

questions, a Wilcoxon Signed Rank Test was performed and it showed that there was a statistically significant positive change or improvement to the test scores of students, z = –2.427, p = 0.015. However, the effect size calculated as r = 0.30, indicating a medium effect size (~~observed to be small~~ r = 0.02). The Wilcoxon test results revealed that the scores of 32 students showed a positive increase, while 14 students showed a decrease in test scores, and there were no changes to the scores of 18 students. Table 3 summarises the test results.

Table 3. Wilcoxon test results for direct quiz questions

Comparison	Type of rank	N	Mean rank	Sum of ranks
DQ_POST – DQ_PRE	Negative ranks	14[a]	23.21	325.00
	Positive Ranks	32[b]	23.63	756.00
	Ties	18[c]		
	Total	64		

Notes: a = DQ_POST < DQ_PRE; b = DQ_POST > DQ_PRE; c = DQ_POST = DQ_PRE

Problem-Based Questions (PBQ) Quiz Scores: Sample Characteristics and Results

Similar to the direct quiz questions, the students were asked to answer five problem-based multiple-choice quiz questions in both the pre-test and post-test. The sample distribution was not normal, as indicated by the normality test. Because of this, the Wilcoxon Signed Rank Test was employed as a nonparametric alternative to the paired-samples t-test to assess the impact of ChatGPT on the five problem-based quiz questions. To assess whether there were any changes in quiz scores when students used ChatGPT as an additional learning tool to answer problem-based multiple-choice quiz questions, a Wilcoxon Signed Rank Test showed a statistically significant positive change in test scores, z = –2.432, p = 0.015. However, the effect size calculated as r = 0.30, indicating a medium effect size (~~observed to be small~~ r = 0.02). The Wilcoxon test results revealed that the scores of 34 students showed a positive increase while 15 students' test scores showed a decrease and there were no changes to the scores of 15 students. Table 4 summarises the test results.

Table 4. Wilcoxon Test results for problem-based quiz questions

Comparison	Type of rank	N	Mean rank	Sum of ranks
PBQ_POST – PBQ_PRE	Negative ranks	15[a]	25.13	377.00
	Positive ranks	34[b]	24.94	848.00
	Ties	15[c]		
	Total	64		

Notes: a = PBQ_POST < PBQ_PRE; b = PBQ_POST > PBQ_PRE; c = PBQ_POST = PBQ_PRE

DISCUSSION

The research findings from this exploratory study suggest that after using ChatGPT students' overall quiz scores improved significantly, with an average rise of 9.69%. From the perspective of Bloom's Taxonomy, this improvement suggests that ChatGPT effectively supports lower-order cognitive processes such as remembering and understanding, and to some extent applying knowledge. However, its limited impact on achieving top grades indicates weaker support for higher-order skills such as evaluation and creation, which are essential in legal education (Anderson & Krathwohl, 2001). Given that the passing threshold at Taylor University is a score of 50 and that a score of 75 is required for an A– grade (with 80 or above being an A), this translates to nearly a two-grade rise. It is important to note that due to the lack of a control group, these observed gains should be interpreted with caution; the findings indicate a strong correlation, but a direct causal link solely attributable to ChatGPT is difficult to establish. Additionally, the use of identical pre-test and post-test questions may have introduced a familiarity effect, potentially inflating post-test performance independently of ChatGPT usage (Campbell & Stanley, 1963; Roediger & Butler, 2011; van Gog & Sweller, 2015; Yang et al., 2021). It was also discovered that the number of students who scored 50 or more on the pre-test was 19 which increased to 33 after using ChatGPT on the post-test, resulting in a 73.7% increase in the number of students passing the quiz (score of 50 or more). The percentage of students who achieved 50 marks increased by 25% after utilising ChatGPT, while the percentage of students who scored 60 and 70 marks increased by 140% and 200%, respectively. However, no students received a score of more than 70 on both the pre-test and post-test. The highest performance remained at the B range (70 marks). The study results revealed that ChatGPT could assist students improve their overall performance to achieve pass to good marks, but it could not help students attain very good or excellent grades. In other words, ChatGPT could not help students improve their performance to achieve a grade of A or A–. This result correlates with other studies conducted earlier at the University of Minnesota Law School, (Choi et al., 2022), the University of Pennsylvania (Terwiesch, 2023), and a study conducted by Perlman (2023) as discussed in the literature review. Table 5 summarises the pre-test and post-test scores.

Table 5. Pre-test and post-test scores

Scores	Frequency (Pre-test)	Frequency (Post-test)	Increase (+) / Decrease (–) (No. of students)	Increase (+) / Decrease (–) (%)
0	1	0	–1	–100
10	2	1	–1	–50
20	12	1	–11	–92
30	15	11	–4	–27
40	15	18	3	20
50	12	15	3	25

(Continued on next page)

Table 5. *(Continued)*

Scores	Frequency (Pre-test)	Frequency (Post-test)	Increase (+) / Decrease (–) (No. of students)	Increase (+) / Decrease (–) (%)
60	5	12	7	140%
70	2	6	4	200%
Total	64	64		

The study findings indicate that ChatGPT's intervention had a small effect size concerning different types of quiz questions (direct and problem-based). Nevertheless, it is noteworthy that a significant percentage of students improved their quiz scores on both question types after using ChatGPT. Specifically, 50% of students showed improvement in direct questions, and 53% showed improvement in problem-based questions. Immediately following the completion of the post-test, all participating students were presented with a separate, voluntary, and anonymous open-ended questionnaire to gather qualitative data on their experience. The questionnaire prompted students with the question: "Please provide any comments or feedback regarding your experience using ChatGPT as a companion for the quiz, focusing on both positive and negative aspects." This intentional collection method was used to gather rich, contextual data to complement the quantitative performance metrics. The qualitative findings further support the quantitative results by highlighting that while ChatGPT enhances accessibility to information, it does not necessarily translate into the deeper conceptual understanding required for higher-level performance. A total of 64 students completed the open-ended feedback questionnaire. The six students' comments presented in Table 6 are selected, illustrative quotes chosen to represent experience using ChatGPT as an aid for the quiz. A thematic analysis of the qualitative responses revealed recurring themes, including perceived usefulness for quick clarification, concerns regarding accuracy, and instances of over-reliance. These themes provide a more systematic understanding of students' experiences rather than isolated anecdotal observations (Braun & Clarke, 2006). These six students' quotes are not intended to be a representative sample of all students, but rather to provide in-depth examples of their experiences using ChatGPT. The respondents' anonymity was maintained throughout the collection and reporting process.

Table 6. Students' responses on using ChatGPT in their quiz.

Student	Feedback
1	I feel the most significant advantage ChatGPT has over traditional methods of studying is its interactive nature. It often provides immediate results, but unlike Google, it isn't a one-sided mode of research. Students are allowed to 'volley' with the AI programming and instantaneously receive feedback. This encourages its users to research further as it eliminates the gruelling, time-consuming process of having to scour the internet, which is often discouraging.

(Continued on next page)

Table 6. *(Continued)*

Student	Feedback
2	ChatGPT is great honestly. It clearly states that it takes its source from various sources which may or may not be accurate, so obviously, you're not supposed to just plagiarise the answers 100%. ChatGPT makes things easier to understand and it creates clarity when some lecture slides are often from textbooks, which can be confusing and lengthy. However, the reliability of slides and textbooks is better.
3	It was much easier to use ChatGPT compared to the lecture notes. I managed to finish my quiz much faster when I used ChatGPT. It's very convenient and easy to use, as I did not have to look for the answers through the slides. However, ChatGPT might not be as reliable as the lecture slides. Using the lecture slides is a safer bet for finding correct answers, but in my opinion, ChatGPT is a faster and more convenient tool to use.
4	The slides were really helpful in having a broad understanding of the situation, but ChatGPT helped explain and understand the exact question, and it was quick to give a response to the question.
5	To be honest this is the first time I used ChatGPT and I'm so impressed with this AI.
6	ChatGPT is reliable for me, I may need to alter the question a little in order for me to get an accurate answer, but it provides me information in seconds and also an explanation to the answers produced, sure there are some flaws in ChatGPT, but I think I can use it as an efficient way of learning and finding information.

The results directly address RQ1 (overall performance improvement) and RQ2 (passing to good grades). The significant mean performance increase (RQ1) aligns with the general positive outlook in the literature, which highlights ChatGPT's potential as a highly accessible and immediate learning companion (Zhai, 2022). This finding corroborates evidence from studies in other domains, such as the finding that ChatGPT could pass law school exams with an average grade of C+ (Terwiesch, 2023). The effectiveness of the tool in improving student performance from the passing to the good grade range (RQ2) suggests that ChatGPT excels at handling the foundational, information-heavy components of the legal quiz, acting as an efficient information retrieval and synthesis tool for basic legal concepts. This aligns with constructivist learning theory, where tools like ChatGPT may function as scaffolding mechanisms that assist learners in building foundational understanding. However, excessive reliance on such scaffolding may hinder the development of independent critical thinking skills (Frank, 2024; Hwang et al., 2020; Tran et al., 2025). However, this study's most critical finding relates to RQ3 (achievement of highest academic grades), where the research found no significant evidence that ChatGPT assists students in attaining the highest academic grades. In other words, since the highest grades are defined by Taylor University as A (80 and above) and A– (75 and above), ChatGPT could not help students improve their performance to achieve these top grades. Notably, even with ChatGPT, none of the 64 students attained an 'A' grade. This finding suggests that while ChatGPT can enhance performance efficiency, it may not facilitate the depth of cognitive processing required for expert-level performance, reinforcing concerns that generative AI may promote surface learning rather than deep learning (Biggs & Tang, 2011). This aligns with prior findings that while AI can handle straightforward queries,

it often falls short on complex, high-order problem solving or may provide superficial answers (Stott & Stott, 2023; Choi et al., 2022). The result suggests that achieving an A-grade likely requires deeper understanding or nuanced application that ChatGPT could not furnish, at least in this instance. The highest performance remained at the B range (70 marks). This is a notable contradiction to the theoretical ceiling of AI performance and provides empirical evidence for the scepticism noted by many educationists (Terwiesch, 2023).

The inability of ChatGPT to significantly elevate performance into the highest academic band (A/A–) can be theorised based on the nature of higher-order legal education. Achieving excellent grades in law often requires more than simple knowledge recall or basic analysis; it demands nuanced judgement, critical evaluation of conflicting legal principles, synthesis of complex, context-dependent case facts, and an empathetic understanding of policy implications. While ChatGPT is powerful for factual queries and structuring responses, it fundamentally lacks the human instructor's nuanced understanding and empathetic insight—a limitation posited in the literature (e.g., Su & Yang, 2023). The top grades are earned by demonstrating these advanced cognitive skills, which a partial level of ChatGPT automation cannot fully replicate or instruct. The findings thus support the literature's caution that over-reliance on ChatGPT might discourage the deep critical thinking essential for mastering intricate legal matters and achieving true expertise.

Crucially, a subset of students demonstrated a negative change in performance in the post-test after using ChatGPT. There were 14 students who scored lower on the direct questions and 15 students scored lower on the problem-based questions. This counter-intuitive finding highlights the complex relationship between AI assistance and student learning, suggesting that the integration of the tool can, for some, be detrimental. This phenomenon may also be explained by cognitive load theory, where the additional effort required to interpret and evaluate AI-generated responses may overload learners, particularly those with weaker foundational knowledge (Sweller, 1988; Skulmowski & Xu, 2022; Twabu, 2025). This negative impact may be attributed to several factors inherent in using generative AI during an assessment. Students may have been misled by plausible but incorrect answers generated by ChatGPT, a known limitation of large language models. Inexperienced users may have spent too much valuable quiz time crafting effective prompts or vetting the AI's responses, leading to overall time mismanagement and missed questions. These students might have over-relied on the AI's answer without cross-referencing their lecture notes, trusting the output over their own foundational knowledge, which could be particularly risky if the AI provided contextually inaccurate legal information.

From a pedagogical perspective, these findings suggest that generative AI tools should be integrated as supplementary learning aids rather than primary instructional tools, particularly in disciplines such as law that require higher-order reasoning and critical analysis. In light of the findings, ChatGPT should be viewed as a reliable and effective study companion for scaffolding foundational knowledge, but not as a replacement for human mentorship in developing advanced legal judgement (RQ4). It works best when

integrated to enrich the overall learning experience by handling repetitive information tasks, thereby freeing up both student and instructor time for high-value activities that foster critical thinking.

This study contributes to the growing body of literature on generative AI in education by providing empirical evidence from a discipline-specific context, highlighting both the performance benefits and pedagogical limitations of ChatGPT in legal education.

CONCLUSION

This study set out to evaluate the efficacy of ChatGPT as a study companion in legal education using a pre-test/post-test design under the guidance of the IDEE framework. The principal contribution of this study is providing one of the first empirical assessments of ChatGPT's impact on legal student performance, moving beyond anecdotal and theoretical discussions. By situating these findings within a discipline-specific context, this study contributes to the emerging literature on generative AI in education by highlighting the differential impact of ChatGPT across levels of cognitive complexity. The results clearly indicate that the use of ChatGPT significantly improves student performance in legal education, particularly helping them achieve passing to good grades. From a pedagogical perspective, this finding suggests that ChatGPT is particularly effective in supporting lower-order cognitive processes such as knowledge acquisition and comprehension but has limited capacity to facilitate higher-order thinking skills such as evaluation and synthesis (Anderson & Krathwohl, 2001). However, the study concludes that ChatGPT's utility diminishes at the highest academic levels, confirming that the critical thinking and nuanced judgement required for achieving A or A-grades remains fundamentally a human domain. This reinforces constructivist perspectives that emphasise the role of human interaction and guided instruction in developing deep learning and critical reasoning (Frank, 2024; Hwang et al., 2020; Tran et al., 2025). ChatGPT should not replace the comprehensive guidance and feedback that human instructors and mentors can provide. It's most effective when used in conjunction with traditional legal education methods to enhance the overall learning experience.

The limitations of this study must be acknowledged. The exploratory nature of the single-group pre-test/post-test design limits the establishment of definitive causal claims, and the small, single-institution sample size affects generalisability. These limitations suggest that the findings should be interpreted as indicative rather than definitive, particularly in relation to causal inference. Furthermore, the use of identical quiz questions introduces a potential learning effect on post-test scores, and the study only measured immediate performance, lacking data on longer-term learning, retention, or the development of higher-order skills like critical thinking. Based on these findings and limitations, future research should adopt more rigorous designs, such as conducting a Randomised Control Trial (RCT) with a control group to isolate the causal effect of ChatGPT, preferably using alternate but equivalent questions in the pre- and post-tests to mitigate the learning effect. Such designs would enable a clearer distinction between performance enhancement and genuine learning

gains. Future research should also explore the impact of using different levels of ChatGPT automation (per the IDEE framework) and employ systematic qualitative investigation, such as in-depth surveys on student engagement or detailed qualitative analysis of student attitudes, to fully explore the companion role (RQ4) beyond immediate quiz scores. Finally, evaluating the tool's efficacy using different, high-stakes legal assessments, such as moot court briefs or essay-based case studies which require deeper critical thinking than multiple-choice quizzes is essential to provide a more comprehensive picture of ChatGPT's utility. From a pedagogical standpoint, these findings underscore the need for educators to design learning environments that integrate AI tools while simultaneously fostering independent critical thinking and reflective learning practices.

In conclusion, ChatGPT should be seen as a powerful, supplementary tool for scaffolding foundational knowledge in law, but not a substitute for the core pedagogical relationship between student and human instructor.

REFERENCES

Ajevski, M., Barker, K., Gilbert, A., Hardie, L., & Ryan, F. (2023). ChatGPT and the future of legal education and practice. *The Law Teacher, 57*, 352–364. https://doi.org/10.1080/03069400.2023.2207426

Alarie, B. (2023, January 2). The rise of the robotic tax analyst. *Tax Notes Federal, 178*, 57.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Longman.

Ball, C. (2023, January 27). ChatGPT proves a mediocre law student. *Ball in Your Court.* https://craigball.net/2023/01/27/chatgpt-proves-a-mediocre-law-student/

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Open University Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin.

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education, 20*(1), 38. https://doi.org/10.1186/s41239-023-00408-3

Choi, J. H., Hickman, K. E., Monahan, A. B., & Schwarcz, D. (2022). ChatGPT goes to law school. *Journal of Legal Education, 71*(3), 387–400. https://doi.org/10.2139/ssrn.4335905

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

Dimitrov, D. M., & Rumrill, P. D. (2003). Pretest-posttest designs and measurement of

change. *Work, 20*(2), 159–165. https://doi.org/10.3233/WOR-2003-00285

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., Al-Busaidi, K. A., 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

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics, 5*(1), 1–4. https://doi.org/10.11648/j.ajtas.20160501.11

Frank, A. (2024). Artificial intelligence as a scaffolding tool for student learning: Implications for higher education. *Frontiers in Education, 9*, 1793940. https://doi.org/10.3389/feduc.2024.1793940

Hargreaves, S. (2023). "Words are flowing out like endless rain into a paper cup": ChatGPT & Law School Assessments. *Legal Education Review, 33*(1), 69–105. https://doi.org/10.53300/001c.83297

Hu, R., & Lin, X. (2025). Incorporating ChatGPT into online discussions in a literacy course: Impact and students' perceptions. *International Journal of Artificial Intelligence (AI) in Teaching and Learning, 1*(1), 1–18. https://doi.org/10.4018/IJAITL.366589

Hwang, G. -J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence, 1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Irfan, M., Murray, L., & Ali, S. (2023). Integration of artificial intelligence in academia: A case study of critical teaching and learning in higher education. *Global Social Sciences Review, 8*(1), 352–364. https://doi.org/10.31703/gssr.2023(VIII-I).32

Javaid, M., Haleem, A., Singh, R. P., Khan, S., & Khan, I. H. (2023). Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system. *BenchCouncil Transactions on Benchmarks, Standards & Evaluations, 3*(2), 1–12. https://doi.org/10.1016/j.tbench.2023.100115

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Keith, M., Keiller, E., Windows-Yule, C., Kings, I., & Robbins, P. (2025). Harnessing generative AI in chemical engineering education: Implementation and evaluation of the large language model ChatGPT v3.5. *Education for Chemical Engineers, 51*, 20–33. https://doi.org/10.1016/j.ece.2025.01.002

Kiryakova, G., & Angelova, N. (2023). ChatGPT: A challenging tool for the university professors in their teaching practice. *Education Sciences, 13*(10), 1056. https://doi.org/10.3390/educsci13101056

Klingensmith, M. W. (2023). Let's talk, ChatGPT: What will the judiciary's future look

like? *Florida Bar Journal, 97*(3), 26–33.
Lee, H. (2024). The rise of ChatGPT: Exploring its potential in medical education. *Anatomical Sciences Education, 17*, 926–931. https://doi.org/10.1002/ase.2270
Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), 410. https://doi.org/10.3390/educsci13040410
Lock, S. (2022, December 5). What is AI chatbot phenomenon ChatGPT and could it replace humans? *The Guardian.* https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans
Miller, M. (n.d.). Artificial intelligence (AI). *Ditch That Textbook.* https://ditchthattextbook.com/ai/
OECD. (2021). *AI and the future of skills, Volume 1: Capabilities and assessments*. OECD Publishing. https://doi.org/10.1787/5ee71f34-en
Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence, 2*, 100033. https://doi.org/10.1016/j.caeai.2021.100033
Oltz, T. P. (2023). ChatGPT, professor of law. *University of Illinois Journal of Law, Technology & Policy, 2023*(1), 207–221.
OpenAI. (n.d.). GPT-4 is OpenAI's most advanced system, producing safer and more useful responses. https://openai.com/product/gpt-4
Perlman, A. M. (2023, March/April). The implications of ChatGPT for legal services and society. *The Practice.* https://clp.law.harvard.edu/knowledge-hub/magazine/issues/generative-ai-in-the-legal-profession/the-implications-of-chatgpt-for-legal-services-and-society/
Riermaier, P. (2023). *ChatGPT and other AI technologies in the study and practice of law*. Penn Carey Law, University of Pennsylvania. https://www.law.upenn.edu/live/news/15538-chatgpt-and-the-law
Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Houghton Mifflin.
Skulmowski, A., & Xu, K. M. (2022). Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. *Educational Psychology Review, 34*(1), 171–196. https://doi.org/10.1007/s10648-021-09624-7
Stokel-Walker, C. (2023). ChatGPT listed as an author on research papers: Many scientists disapprove. *Nature, 613*(7945), 620–621. https://doi.org/10.1038/d41586-023-00107-z
Stott, F. A., & Stott, D. M. (2023). A perspective on the use of ChatGPT in tax education. In T. G. Calderon (Ed.), *Advances in accounting education: Teaching and curriculum innovation*s (Vol. 27, pp. 145–153). Emerald Publishing Limited. https://doi.org/10.1108/S1085-462220230000027007
Su, J., & Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, 20965311231168423.
Sullivan, W. M., Colby, A., Wegner, J. W., Bond, L., & Shulman, L. S. (2007). *Educating*

lawyers: Preparation for the profession of law. Jossey-Bass.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Taylor, S. T. (2023). Taylor's perspective: Pros & cons of ChatGPT for law firms according to ... ChatGPT. *Of Counsel, 42*(4), 3–4.

Terwiesch, C. (2023). *Would ChatGPT get a Wharton MBA? A prediction based on its performance in the operations management course.* https://mackinstitute.wharton.upenn.edu/wp-content/uploads/2023/01/Christian-Terwiesch-Chat-GTP.pdf

Tran, T., Nguyen, H., Pham, L., & Le, D. (2025). Artificial intelligence as a "more knowledgeable other": Exploring generative AI in learning contexts. *npj Digital Medicine*. Advance online publication. https://doi.org/10.1038/s41746-025-01823-8

Twabu, T. (2025). Artificial intelligence and cognitive load: Implications for learning and instruction. *Discover Education, 4*, Article 592. https://doi.org/10.1007/s44217-025-00592-6

van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: The testing effect decreases with increased complexity. *Educational Psychology Review, 27*(2), 247–264. https://doi.org/10.1007/s10648-015-9310-x

Ventayen, R. J. M. (2023). ChatGPT by OpenAI: Students' viewpoint on cheating using artificial intelligence-based application. *Social Science Research Network*. https://doi.org/10.2139/ssrn.4361548

Ward, S. F. (2023). STUDY AIDS: Some law schools already are using ChatGPT to teach legal research and writing. *ABA Journal, 109*(3), 38–39.

Weiler, J. H. H. (2023). Editorial: ChatGPT and law exams; On my way in IV: "Aren't you exclusive?!" On the pros and cons of writing letters of reference for only one candidate in an academic hiring process; In this issue; In this issue - reviews. *European Journal of International Law, 34*(2), 281–290. https://doi.org/10.1093/ejil/chad029

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin, 147*(4), 399–435. https://doi.org/10.1037/bul0000309

Yin, J., Goh, T. -T., Yang, B., & Xiaobin, Y. (2021). Conversation technology with micro-learning: The impact of chatbot-based learning on students' learning motivation and performance. *Journal of Educational Computing Research, 59*(1), 154–177. https://doi.org/10.1177/0735633120952067

Zhai, X. (2021). Practices and theories: How can machine learning assist in innovative assessment practices in science education? *Journal of Science Education and Technology, 30*(2), 139–149. https://doi.org/10.1007/s10956-021-09901-8

Zhai, X. (2022). ChatGPT user experience: Implications for education. *Social Science Research Network*. https://doi.org/10.2139/ssrn.4312418