A Formative Assessment Approach to Examine Cognitive and Attitudinal Effects of AI-based LLM Use among Undergraduate Engineering Students

C. Sivapragasam[1*], M. Vasudevan[2*], N. Natarajan[3], P. Saravanan[4]
[1]Department of Civil Engineering, Kalasalingam Academy of Research and Education, Srivilliputhur, Tamil Nadu, India
[2]Department of Agricultural Engineering, Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India
[3]Department of Civil Engineering, Dr. Mahalingam College of Engineering and Technology, Pollachi, Tamil Nadu, India
[4]Department of Civil Engineering, Dr. Ambedkar Institute of Technology, Bengaluru, Karnataka, India
[1]sivapragasam@klu.ac.in, [2]devamv@gmail.com, [3]itsrajan2002@yahoo.co.in, [4]p.saravanan.me.1986@gmail.com

***Abstract*—The pace of integrating artificial intelligence (AI)-based large language models (LLMs) into education and research has grown exponentially in recent years. Being the key stakeholders, the nature and extent of dependence on AI tools among the educators, researchers, and students vary considerably across multiple factors, posing critical challenges in systematically evaluating and interpreting their impacts. The present study proposes a formative assessment approach to compare the comprehension level of selected fundamental engineering concepts for a group of eight students through (i) an offline in-person proctored tests and (ii) online feedback surveys. The tests comprise two sets of multiple-choice and short-answer questions with increasing cognitive levels, while the use of AI tools was permitted for the second set. This is followed by a feedback survey to capture the nature of responses to questions with and without AI assistance. The results show a growing dependency on AI tools for answering conceptual and analytical questions compared to factual and recall-type questions. The usage of AI tools showed a three-fold hike in the overall performance in the open-book test (78.75%) compared to the proctored assessment (26.25%). The observed patterns of AI use indicate a shift toward more methodological searches compared to random ones. The study recommends that students should first build strong conceptual foundations through conventional learning, and AI use for assignments or projects should be discouraged at least until the second year to ensure they develop independent thinking skills. Post-assessment follow-ups with mentoring has to be adopted, with attention to their deeper implications for behavioral traits and intellectual responsiveness.**

***Keywords*—AI in education; Digital use patterns; Impacts of LLM; Formative assessment; Cognitive Levels**

***ICTIEE* Track—Assessment, Feedback, Learning outcomes**
ICTIEE Sub-Track: Measuring higher order thinking and critical thinking.

I. INTRODUCTION

The technological innovation is driving the higher education sector towards faster reformations giving rise to increased levels of expectations about performance assessment by utilizing advanced tools for flexible and adaptive learning environments. In India, the implementation of national education policy (NEP) and national credit transfer framework (NCrF) are among the chief contemporary transformations acknowledging the vibrant development of learning platforms. In addition, the integration of information communication tools (ICT) has been widely practiced as an effective methodology for adaptive learning (Sivapragasam & Natarajan, 2023, Sivapragasam et al., 2024). At the latest stage of using these ICT tools, the integration of Artificial Intelligence (AI) tools, particularly large language models (LLMs), has emerged with the potential to reshape teaching, learning, and assessment practices (Klymkowsky & Cooper, 2024). The ways and means by which these LLM tools have been customized and made accessible in the academic world indicate not merely a technological evolution but a paradigm shift in the educational transaction and learning process (Akolekar et al., 2025). In general, the generative AI tools based on deep learning (DL) and natural language processing (NLP) techniques are increasing their capacities for performing diverse and complex tasks, including content generation as well as analytical problem-solving (Ooi et al., 2025).

The academic practices of undergraduate engineering students typically demand deeper levels of conceptual understanding of fundamental subjects such as fluid mechanics, thermodynamics, and engineering design and analysis. Engineering fluid mechanics, for instance, involves fundamental concepts from physics that are practically oriented and applied in various flow systems and hydraulic or pneumatic

[1]Dr. C Sivapragasam, [2]Dr M Vasudevan
[1]Kalasalingam Academy of Research and Education, Srivilliputhur, Tamil Nadu, India
[2]Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India
[1]sivapragasam@klu.ac.in, [2]devamv@gmail.com

machinery. A lack of proper comprehension of these concepts makes it difficult to solve analytical problems or design fluid flow components. Many students struggle to understand the basic principles and their related applications, which often reflects in their poor academic performance and difficulty in passing competitive examinations based on core subject knowledge (Brown, 2018; Belim et al., 2025). Conventional assessment methods typically include written tests with multiple-choice, descriptive, and analytical problem-solving questions, as well as laboratory-based experiments. Quite often, they fail to measure some of crucial skills needed by modern engineers, including critical thinking, creativity, and practical wisdom. To address the limitations in these methods, several researchers have proposed innovative approaches and assessment tools for evaluating the student performance in different contexts (Sundar et al., 2020; Natarajan et al., 2020; Beneroso & Robinson, 2021; Raje & Tamilselvi, 2024; Cossu et al., 2024; Li & Cheung, 2025). Among these, the formative assessment provides an opportunity for *en route* evaluation and corrections. Compared to the summative evaluation that primarily measures achievement, formative assessment provides feedback, reflection, and iterative improvement (Cole and Spence, 2012). Thus, it provides a natural framework for simultaneously analyzing the cognitive outcomes and attitudinal responses associated with the use of AI tools for academic preparations.

In addition to evaluating the performance of the students in terms of cognitive skills, understanding their emotional responses and attitudinal trends with the increased AI usage is certainly a crucial question to be answered (Pawar et al., 2025). Although widely perceived as an opportunity for boosting confidence and problem-solving ability, an increasing trend in the dependency of AI tools by the students must be viewed with anxiety and doubt about their personal, social, and environmental implications (Farrokhnia, 2024; Suhonen, 2025). Concerns also remain about the cognitive consequences of over-reliance, as well as issues of authenticity and academic integrity. Zheng et al. (2023) reported the weakness of ChatGPT in answering open-domain questions, identifying challenges related to comprehension, factuality, specificity, and inference. Factuality of the results is a significant issue linked to deficiencies in recalling and memorizing knowledge. The authors suggested that reliability of LLMs can be improved by integrating external knowledge and recall cues, but the study did not address the psychological aspects of the students. Johnson et al. (2023) assessed performance of ChatGPT on 284 medical questions across 17 specialties and demonstrated high median accuracy (5.5/6) and completeness (3/3). They found that the limitations persisted, particularly with difficult questions by analyzing re-querying improved responses.

Amaro et al. (2023) found that while ChatGPT occasionally produced unreliable outputs, computer science students continued to rely on it, acknowledging the need for human intervention for confirmation. When Ngo et al. (2024) tested ChatGPT 3.5 in generating 60 multiple-choice questions with explanations, only 32% were fully accurate, while 25% contained misleading content. This confirmed that while ChatGPT could assist in generating extensive study materials, human editing cannot be avoided to make proper use of the acquired information. This is particularly observed as crucial in medical studies, where the AI model produced erroneous interpretations in complex cases such as hyponatremia (Berend et al., 2025). The inconsistency and algorithmic bias in LLMs thus may confuse students regarding their reliability rather than enriching their knowledge. It is important to recognize that emotional behaviors such as anxiety, guilt, or overconfidence can directly affect motivation and performance of students in academics. A structured mentoring system is therefore essential to guide the students, provide constructive feedback, and encourage critical engagement with AI tools in their academic or non-academic preparations.

In engineering education, where problem-solving requires both conceptual clarity and emotional resilience, a critical gap remains in understanding how formative assessment can be leveraged to evaluate and guide students' use of AI tools toward constructive learning practices (Hudesman et al. 2013; Fütterer et al., 2023; Cotton et al., 2024; Ramprakash et al., 2024). As the AI tools are prone to make significant errors with many *wrong-as-right* (type I) errors, great risk is posed to the students if they are not strong in the fundamental concepts of the subjects. In this aspect, the Revied Bloom's Taxonomy (RBT) facilitates a systematic assessment of intellectual progress to higher-order thinking skills (HOTS) (Na et al., 2021; Qadar et al., 2025; Li et al., 2025; Xiao et al., 2025). Most of the recent studies highlight the scope of using AI tools for facilitating tailor-made and quickly adaptive means of learning including formative assessment, while the long-term implications of their immature immersion with the academic practices requires careful considerations (Zhai and Nehm, 2023; Hopfenbeck et al., 2023; Li et al., 2023). The main objective of the present study is to evaluate the influence of AI tools on the behavioral responsiveness of undergraduate engineering students while answering concept-based and analytical problem-solving questions across varying cognitive levels. Two types of formative tasks: closed-book and open-book tests, were employed to compare students' ability in answering fundamental questions aligned with the RBT levels with and without dependence on AI tools. Additionally, the extent and typical nature of AI use were analyzed through a feedback survey in the form of a three-level hierarchical questionnaire exploring the patterns of search, storage, and data retrieval activities for academic purposes. By placing emphasis on learning the basic concepts in the conventional way, the study brings deeper insights into cognitive, behavioral, emotional, and ethical impacts of early adoption of AI tools by the students for their academic activities.

II. Materials and Methods

A. Background

The study was conducted with the second-year batch of undergraduate students from the department of agricultural engineering having a class strength of 57. The subject of focus was *Fluid Mechanics* as part of their third semester curriculum owing to its wide range of applications in water flow, storage, and irrigation systems. For this research experiment, only selected fundamental concepts from fluid mechanics were chosen to design the formative assessment questions. These questions were prepared in multiple sets by several faculty members who had previously taught the subject on multiple occasions, thereby ensuring validity and consistency. The student participants were selected from the current cohort with equal gender representation (1:1), based on their academic background (class grades A, B and C) and their willingness to share details of their academic preparation. As part of the ethical clearance, the students were informed early about the test pattern and assessment conditions. All participants attended both assessments as per the planned schedule and demonstrated their conceptual understanding and problem-solving skills during the evaluation.

B. Formative Assessment Approach

The conceptual understanding and analytical skills of the students in the subject Fluid Mechanics were assessed using a question bank prepared by faculty members. The questions focused on fundamental topics such as fluid properties, hydrostatic pressure, and fluid flow, and were designed with progressively increasing levels of cognitive demand based on the RBT scheme. Specifically, the first five prominent levels of RBT were considered for the assessment: Remember–Factual (*R/F*), Understand–Conceptual (*U/C*), Apply–Conceptual (*Ap/C*), Apply–Procedural (*Ap/P*), and Analyze–Conceptual (*An/C*). The assessment included multiple-choice and short-answer type questions, with detailed contextual information in the form of scenarios also incorporated wherever necessary to enhance clarity and practical relevance of the questions (Kadiyala et al., 2017). Two sets of question papers were developed for the (i) closed-book test and (ii) open-book test after ensuring content alignment and validating difficulty levels through expert review by faculty members (Refer Tables S1 and S2). A summary of the metadata for the question papers is presented in Table 1.

An online survey was conducted using a customized Google Form to collect feedback from the participants regarding the tests (Table S3). Unlike the fixed-time assessments, the survey responses were collected after allowing sufficient time for reflection and recollection. Based on prior experience, it was observed that providing at least an overnight period enabled students to give more assertive and comprehensive responses. Analysis of the response sheet indicated that most students preferred to complete the Google Form during the morning hours, reflecting their routine pattern of engaging with the internet for academic activities. As part of the assessment, students were informed about the purpose and importance of the survey to ensure sincerity in their responses and to gather deeper insights into their learning characteristics. The metadata of the online survey is presented in Table 2.

A novel formative assessment approach was adopted in this study, wherein the students first participated in a closed-book classroom examination, followed by an open-book examination that permitted the exceptional use of AI tools accessible through their mobile devices. To complement these assessments, an online feedback survey was also conducted to gather participants' reflections on their performance, including their levels of preparation, answering strategies, and the perceived implications of AI tool usage. Though the survey formed part of the formative assessment, it was conducted online on the following day to ensure more accurate post-exam reflections. As the study aims to compare the students' performance between two formats of assessments, the existence of mean-based variability as well as the significance of uniformity (or non-uniformity) in the response pattern are to be investigated. In order to extract the nature of association between the study variables (i.e., the cognitive levels, students' attempts and performance), a few statistical measures such as descriptive statistics and paired t-tests were performed using the Data Analysis Toolpack of Microsoft Excel Spreadsheet. A flowchart summarizing the implemented methodology is presented in Fig. 1.

III. Results and discussion

A. Types of Assessment and the Nature of Responses

The results from the closed-book and open-book tests revealed prominent differences in students' learning patterns when engaging with conceptual questions in the engineering subject Fluid Mechanics. In the closed-book test, most students spent additional time but used relatively less paper space, particularly for higher-order questions. Their performance on factual and conceptual questions was notably weaker compared to procedural ones. Many students showed limited inclination to provide proper justifications for their answers, especially in analytical problems where additional marks were allocated for explanation. Higher-order thinking-type (HOT) questions demanded significant effort in terms of critical thinking and scenario revisiting. In contrast, the open-book test – despite being based on a similarly structured question paper – was completed in a shorter span of time, irrespective of the variations in the RBT level of the questions. The allowance to use AI tools superficially demonstrated an improved performance by the students in the open-book test. A summary of the comparative performance across the two types of tests in terms of their scores is presented in Table 3.

Closed-book performance data indicated that concept-based multiple-choice questions were poorly attempted (Table 3). Analytical problems were partially attempted, often without sufficient justification. The average percentage of marks obtained across the chosen RBT levels was 23.71, while the average marks across students was slightly higher at 26.25. The least attempted questions were of the *R/F* (81.25% un-attempted) and *Ap/P* (62.5% un-attempted) types. By contrast, the performance in the open-book test was considerably enhanced with AI support: the average marks across RBT levels rose to 77.46%, and the average across students increased to 78.75%. This situation poses a critical lack of clarity in the fundamental concepts that must be taught in a conventional way (not necessarily with AI tools). The huge difference in their scores also indicate their quick acceptance of AI search results as to be true without being able to crosscheck and verify the answers with their own cognitive skills. At this stage, therefore, it is more important for the students to strengthen their understanding of basic concepts and their applications through cognitive assignments and assessments, rather than relying on virtual assistants for quick and easy solutions.

Despite the improved scores, several anomalies were observed. Incorrect answers persisted, often due to misinterpretation or improper application of AI-generated outputs. For example, a typical *Ap/P*-level question on calculating hydrostatic pressure at 1 m depth of water was incorrectly answered by many students due to unit mismanagement, even though the numerical value was correct. Similarly, for an *Ap/P*-level question on computing the average velocity in a pipe using the continuity equation, students copied ChatGPT-generated elaborations beyond the required steps (up to Reynolds number computation) and overlooked the mark split-up specified in the question. Another challenge emerged in image-based questions. AI tools failed to interpret visual inputs effectively: for an *Ap/C*-level question for interpreting the types of flow patterns from three images (a rotating sprinkler head, a smooth flowing river, and a sloping pipeline), students' answers were vague and non-specific, typical to the AI language. Likewise, in an *An/C*-type image-based scenario, it appears that the students tempted to rely solely on AI-generated image interpretations rather than studying the contextual description. These evidences of anomalies certainly indicate the challenges faced by the students in answering the formative assessment because of their overdependence of AI tools.

B. Use of AI varies with Cognitive Levels of Questions

When comparing the cognitive levels of the questions, variations were observed in students' approaches to information retrieval and problem-solving (Fig. 2). For factual-remember (*R/F*) type questions, most students (62.5%) relied on their own memory of familiar concepts, while 25% preferred keyword-based searches using AI tools. A smaller proportion (12.5%) directly searched the entire question in the AI chatbot. The ChatGPT, accessed primarily through mobile applications, emerged as the most preferred tool for this purpose. For conceptual-understanding (*U/C*) type questions, the feedback indicated that the predominant methods were "*own memory*" and "*direct AI search*," both reported by 37.5% of students. In contrast, keyword searches and random guesses were less common (12.5% each). Notably, answers in this category often lacked specificity and brevity. Consequently, the length of responses in the open-book test scripts was noticeably greater compared to the closed-book test, reflecting reliance on AI-generated elaborations. In the case of apply-procedural-analytical (*Ap/P* and *An/C*) type questions, "*direct AI search*" and "*random guessing*" were the most common strategies, each reported by 25% of students (Fig. 2).

The significance of variations between the students' performance between the two formats of assessment were further investigated using simple statistical measures. The descriptive statistics for the closed-book test results indicated that the difference between the median and mean is highest closer for *U/C*-type questions whereas the standard deviation and variance were higher for *An/C*-type questions (Table 4). This is attributed to the similarity in the number of students attempting the same RBT levels. In case of open-book test results, the mean and median are closer for almost all RBT-levels with high percentage values, while the standard deviation is highest for *An/C*-type questions (Fig. 3). Further, paired t-test results indicated that there is a statistically significant difference between the students' attempting HOT questions and their corresponding performances in closed-book test format (Table 5). This means that the students attempted questions need not give them proportionate marks, indicating their difficulty to score the HOT questions. In the case of open-book test results, the one-tailed test proposed significance in one direction (i.e., marks > attempts), but the two-tailed test does not confirm the significance to be strong. The inferences from the statistical analysis provides further insights to the nature of complexity and the level of confidence in predicting the students' performance using formative assessments.

The feedback based on the closed-book examination revealed that immediate preparation and the remembrance of previous studies helped only one-third of the participants (Figure 2). This indicates the need for improving the understanding of basic concepts in the subject for answering the HOT questions. One common observation was the inherent proportion between the length of answers, mark distribution and mark split-up mentioned in the question paper. Though the answers in the closed-book test indicate the awareness level of the students on the concepts presented in relation to the given scenario by writing something relevant, the answers were not precise. During invigilation, it was observed that most of the students were busy copying the responses from ChatGPT without trying to modify them according to the question, thus increasing the paper space used. Even for questions with lesser marks, the answers were quite lengthy. A major part of the time was consumed by students in searching with keywords and phrases compared to the time spent on customizing the results and writing the answers on paper. It was found that, if exclusive allowance is provided, a large majority (75%) of the students preferred AI tools for answering the questions (Fig. 2). The way in which the students answered the open-book test questions can be corroborated by the fact that their adaptation to regular and frequent use of AI tools was mostly disrupting them to their learning process rather than supporting them.

About 50% of the students considered the search results as correct answers and directly wrote them, while 25% of the students used the AI search results as an aid for checking their memory and conceptual understanding (Fig. 2). This is an alarming situation where the accuracy of representing technical facts and concepts by AI tools can easily misdirect students into blindly believing and relying on them for everything. As mentioned in many earlier reports, the AI is quite prone to provide erratic answers which requires thorough crosscheck before accepting first hand (Fütterer et al., 2023; Ooi et al., 2025). For example, we found that ChatGPT made an error on approximating the value of a dependent parameter (S) from a non-linear equation with another constant (Ks), where the negative sign was omitted, give an absurd result which is meaningless (Fig. 4). Considering the deeper significance of such events, a serious question arises: *can we allow the students get misled by the erroneous results shown by the LLMs, or can we focus on enhancing their conceptual knowledge by which they will be able to decipher the prompted answer as right or wrong?*

One may argue that as the back-end algorithms get frequently updated with high-end computational powers, AI search results may become more easily acceptable and normalized as user-friendly and handy tools for students. It was observed that nearly 62.5% of the participants preferred to scan the questions without even attempting to type them, considering the time factor (Fig. 2). The results reveal that the impact of using AI tools for academic presentations is multi-fold among students, even in assessments, with an increasing trend toward quickly solving questions of higher cognitive levels. This also highlights certain patterns in handling AI tools, mostly developed through regular and repeated usage in general daily affairs. In this context, the reflections provided by the students on AI as post-test feedback are to be taken as important data revealing their attitude and awareness regarding the use of AI tools for academic purposes.

C. Behavioural Patterns and Post-Test Reflections on AI Tools Usage

The permission to use mobile phones during an examination can be viewed as an extreme scenario where the free use of AI tools can be directly visualized and the trends of their influence on students' overall behaviour can be evaluated. The second part of the feedback survey aimed at evaluating the behavioural patterns of AI usage by the participants, with 50% of the questions devoted to this category. The reflective-type questions were well received, and most of the students (75% in total) acknowledged the detrimental effects of AI on their thinking and processing capabilities (Fig. 5). Their responses attributed AI usage to reduced efforts in thinking about small tasks, recalling known facts, and identifying steps in problem-solving. Although 25% of the students claimed they were '*not much dependent*' on AI tools, 37.5% acknowledged that AI tools are helpful for complex/tough questions, while 50% believed in using AI tools irrespective of the nature of the questions. When asked what prompted them to use AI tools in examinations, about 50% of the students cited lack of clarity as well as the toughness of the questions, while 25% admitted they were tempted to save time (Fig. 5). Considering the constraints of using AI tools in fixed-time examinations, students were equally divided in their opinion: some felt they had enough time to answer after making online searches, while others believed it did not make much difference.

While the purpose of conducting an AI-based open test was purely contextual to the present study, we also assessed students' attitudes toward the ethical concerns of such examinations. About 62.5% of the students expressed that permission to answer directly from ChatGPT undermines the very purpose of the examination. However, the remaining 37.5% agreed that AI tools could serve as partial support for remembering basic facts and concepts (Figure 4). On a similar question about using AI tools in all examinations, 87.5% rejected the idea, while 12.5% suggested that simultaneous learning while answering the questions could help in understanding the concepts better. Though this may appear as an appreciable way to improve attentive learning, the possibility of conducting the entire formative assessments as open-book with free access to AI tools does not seem advisable based on the limited scope of the present study.

D. Addressing Emotional Responses for Improved Performance

The emergence of freely accessible AI tools with user-friendly features has generated mixed emotional responses among undergraduate students in conceptualizing the fundamentals of fluid mechanics. The quick availability of simplified yet detailed answers to search queries makes these tools both attractive and addictive to students. Based on the exercise, the students' responses were categorized into three types: preparation, performance, and post-test reflections. Most students reported using AI tools for enhanced conceptual understanding, immediate feedback, and improved problem-solving skills. At the same time, concerns were raised regarding misuse, overdependence, and uncertainty about academic integrity. When asked about the environmental implications of large language models (LLMs), such as water and carbon footprints based on the searching usage and data storage in servers at data centers (Bhaskar and Seth, 2024; Jegham et al., 2025), most students were unaware of these hidden impacts.

Considering the mixed emotional responses toward AI usage, the feedback analysis revealed important behavioural patterns and growing concerns about their implications on academic performance. In this context, a well-defined mentoring system is proposed as a critical intervention to address these issues and to transform the use of AI into a constructive learning experience that enhances formative assessment outcomes. The mentoring framework is conceived in two steps: (i) group interaction (among the students) through scenario-based discussions and (ii) personal interaction (between the student and staff) through reflections on feedback. This approach provides students with greater opportunities to take a personalized approach in addressing their academic challenges. Scheduling a mentoring hour on the day following the assessment could further help students overcome emotional barriers and better comprehend complex concepts while engaging in formative assessments.

It was observed that demonstrating the proper use of AI tools and their appropriate interpretation by the faculty can encourage students to ask metacognitive questions such as: "*What do I understand now*?", "*Why did the AI provide this step*?", and "*How could I verify this independently?*" It is equally important to discuss "*why AI might be wrong*" or "*how to cross-check with first principles*" in order to transform cognitive overdependence into manageable curiosity. When students encounter discrepancies between their own problem-solving approaches and AI-generated solutions, faculty intervention becomes essential to support their efforts and build confidence in critical comparison skills.

Another important concern is the potential sense of unfairness when some students use AI tools extensively while others do not. A recurring ethical issue raised was the guilt associated with '*cheating*' due to excessive reliance on AI during formative assessments. This moral conflict can diminish motivation and create negative emotional associations with learning. Mentoring provides a valuable space for open dialogue on ethical boundaries, responsible use, and the distinction between learning aids and academic dishonesty. Peer monitoring can also normalize responsible practices, reduce competitive anxiety, and foster collective responsibility in the effective use of AI for better performance in formative assessments. We feel that existing academic environment cannot easily address the root causes of the AI usage-damage paradigm merely by putting regulations and enforcements. The academic community has to be sufficiently aware of the all-round dangers of AI usage and should be vigilantly thoughtful in introducing AI tools in the education structure. It is strongly recommended that students up to their third year of undergraduate study be restricted from using AI tools and instead be encouraged to deeply strengthen the fundamental concepts and practice independent problem-solving skills rather than relying on AI-generated outputs. Though the present study focused only on single-subject, single-campus scenario, the key factors are quite similar (and in fact, much more interactive) in large-scale scenarios. We anticipate more such studies should come addressing the root causes and supporting the academic system to retain its core values for humanity.

CONCLUSION

The present study addresses the cognitive and attitudinal dimensions of undergraduate students' engagement with AI tools during formative assessments in technical engineering subjects. A comparative formative assessment is proposed in this study by conducting closed-book and open-book tests for the subject fluid mechanics, followed by a three-level feedback survey. The questions in the tests were framed with an increasing order of cognitive dimensions and the performance of the students with and without use of AI tools were assessed. The results indicate an overdependency trend on AI-based LLMs towards high-order type questions, while conceptual questions were mostly addressed by recollection and reiteration. The nature of using AI tools were significantly favoured by the complexity of questions, lack of clear understanding of the concepts and individual digital behavioural pattern. Most of the students expressed serious emotional concerns on lacking awareness about their deeper cognitive and wider environmental implications. Based on the experience and observed trends, it is strongly recommended that students be systematically trained and empowered in open and critical thinking rather than encouraging the use of AI-enabled tools for solving assignments or for completing projects, at least until second year of their undergraduate studies. A strategic mentoring plan with collaborative interactions is proposed to address the multi-faceted emotional concerns of the students and support them for transforming the AI tools usage as constructive learning experience.

REFERENCES

Akolekar, H., Jhamnani, P., Kumar, V., Tailor, V., Pote, A., Meena, A., & Kumar, D. (2025). The role of generative AI tools in shaping mechanical engineering education from an undergraduate perspective. Scientific Reports, 15(1), 9214. https://doi.org/10.1038/s41598-025-93871-z

Amaro, I., Della Greca, A., Francese, R., Tortora, G., & Tucci, C. (2023). AI Unreliable Answers: A Case Study on ChatGPT. In: Degen, H., Ntoa, S. (eds) Artificial Intelligence in HCI. HCII 2023. Lecture Notes in Computer Science (14051), Springer, Cham. https://doi.org/10.1007/978-3-031-35894-4_2.

Bhaskar, P., & Seth, N. (2024). Environment and sustainability development: A ChatGPT perspective. In Applied Data Science and Smart Systems (pp. 54-62). CRC Press. https://dx.doi.org/10.1201/9781003471059-8

Belim, P., Bhatt, N., Lathigara, A., & Durani, H. (2025). Enhancing Level of Pedagogy for Engineering Students Through Generative AI. Journal of Engineering Education Transformations, 463-470. https://doi.org/10.16920/jeet/2025/v38is2/25057

Beneroso, D., & Robinson, J. (2021). A tool for assessing and providing personalised formative feedback at scale within a second in engineering courses. Education for Chemical Engineers, 36, 38-45. https://doi.org/10.1016/j.ece.2021.02.002

Berend, K., Duits, A., Gans, O.B. (2025) Challenging cases of hyponatremia incorrectly interpreted by Chat GPT. BMC Medical education, 25: 751. https://doi.org/10.1186/s12909-025-07235-2

Brown, A. (2018). Engaging students as partners in developing online learning and feedback activities for first-year fluid mechanics. European Journal of Engineering Education, 43(1), 26-39. https://doi.org/10.1080/03043797.2016.1232372

Cole, J. S., & Spence, S. W. (2012). Using continuous assessment to promote student engagement in a large class. European Journal of Engineering Education, 37(5), 508-525. https://doi.org/10.1080/03043797.2012.719002

Cossu, R., Awidi, I., & Nagy, J. (2024). Critical thinking activities in fluid mechanics–A case study for enhanced student learning and performance. Education for Chemical Engineers, 46, 35-42. https://doi.org/10.1016/j.ece.2023.10.004

Cotton, D. R., 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

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations in Education and Teaching International, 61(3), 460-474. https://doi.org/10.1080/14703297.2023.2195846

Fütterer, T., Fischer, C., Alekseeva, A., Chen, X., Tate, T., Warschauer, M., & Gerjets, P. (2023). ChatGPT in education: global reactions to AI innovations. Scientific reports, 13(1), 15310. https://doi.org/10.1038/s41598-023-42227-6

Hopfenbeck, T. N., Zhang, Z., Sun, S. Z., Robertson, P., & McGrane, J. A. (2023). Challenges and opportunities for classroom-based formative assessment and AI: a perspective article. In Frontiers in Education (Vol. 8, p. 1270700). Frontiers Media SA. https://doi.org/10.3389/feduc.2023.1270700

Hudesman, J., Crosby, S., Flugman, B., Issac, S., Everson, H., & Clay, D. B. (2013). Using formative assessment and metacognition to improve student achievement. Journal of Developmental Education, 37(1), 2. https://files.eric.ed.gov/fulltext/EJ1067283.pdf

Jegham, N., Abdelatti, M., Koh, C. Y., Elmoubarki, L., & Hendawi, A. (2025). How hungry is ai? benchmarking energy, water, and carbon footprint of llm inference. arXiv preprint arXiv:2505.09598. https://doi.org/10.48550/arXiv.2505.09598

Johnson, D., Goodman, R., Patrinely, J., Stone, C., Zimmerman, E. et al. (2023) Asessing the accuracy and reliability of AI-generated medical responses: An evaluation of the Chat-GPT model. Nature Portfolio, https://doi.org/10.21203/rs.3.rs-2566942/v1.

Kadiyala, S., Gavini, S., Kumar, D. S., Kiranmayi, V., & Rao, P. N. S. (2017). Applying blooms taxonomy in framing MCQs: An innovative method for formative assessment in medical students. Journal of Dr. YSR University of Health Sciences, 6(2), 86-91. https://doi.org/10.4103/2277-8632.208010

Klymkowsky, M., & Cooper, M. M. (2024). The end of multiple choice tests: using AI to enhance assessment. arXiv Preprint arXiv:2406.07481. https://doi.org/10.48550/arXiv.2406.07481

Li, T., Reigh, E., He, P., & Adah Miller, E. (2023). Can we and should we use artificial intelligence for formative assessment in science. Journal of Research in Science Teaching, 60(6), 1385-1389. https://ui.adsabs.harvard.edu/link_gateway/2023JRScT..60.1385L/doi:10.1002/tea.21867

Li, R., Li, M., & Qiao, W. (2025). Engineering students' use of large language model tools: An empirical study based on a survey of students from 12 universities. Education Sciences, 15(3), 280. https://doi.org/10.3390/educsci15030280

Li, X., & Cheung, S. C. (2025). A learning-centred computational fluid dynamics course for undergraduate engineering students. International Journal of Mechanical Engineering Education, 53(2), 256-276. https://doi.org/10.1177/03064190231224334

Na, S. J., Ji, Y. G., & Lee, D. H. (2021). Application of Bloom's taxonomy to formative assessment in real-time online classes in Korea. Korean journal of medical education, 33(3), 191. https://doi.org/10.3946/kjme.2021.199

Natarajan, N., Varaprasad, A., & Vasudevan, M. (2020) An Efficient and Simplified Computer Program to Estimate the Infiltration Index. Indian Journal of Geo-Marine Sciences, 49(06), 965-969. http://nopr.niscpr.res.in/handle/123456789/54942

Ngo, A., Gupta, S., Perrine, O., Reddy, R., Ershadi, S., & Remick, D. (2024) ChatGPT 3.5 fails to write appropriate multiple choice practice exam questions. Academic Pathology, 11(1), 100099. https://doi.org/10.1016/j.acpath.2023.100099

Ooi, K. B., Tan, G. W. H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., & Wong, L. W. (2025). The potential of generative artificial intelligence across disciplines: Perspectives and future directions. Journal of Computer Information Systems, 65(1), 76-107. https://doi.org/10.1080/08874417.2023.2261010

Pawar, M., Dhotare, V., Urunkar, N., Andalgavkarkulkarni, Y., Shahane, D., & Pawar, A. S. (2025). Comparative Impact of AI and Search Technologies on Outcome-Based Learning in Engineering Education. Journal of Engineering Education Transformations, 591-598. https://doi.org/10.16920/jeet/2025/v38is2/25073

Qadar, R., Syam, M., & Mahdiannur, M. A. (2025). Analyzing high school physics teachers' understanding of cognitive process and knowledge dimensions in assessment design using the revised Bloom's taxonomy. Discover Education, 4(1), 387. https://doi.org/10.1007/s44217-025-00807-w

Raje, M. S., & Tamilselvi, A. (2024). Gamified formative assessments for enhanced engagement of engineering English learners. Journal of Engineering Education Transformations, 500-507. https://journaleet.in/index.php/jeet/article/view/2451

Ramprakash, B., Nithyakala, G., Bhumika, K., & Avanthika, S. (2024). Comparing traditional instructional methods to ChatGPT: A comprehensive analysis. Journal of Engineering Education Transformations, 612-620. https://journaleet.in/index.php/jeet/article/view/2477

Sivapragasam, C., Dargar, S. K., & Natarajan, N. (2024). Enhancing Engineering Education Through Pedagogical Change: Application to Abstract. Journal of Engineering Education Transformations, 826-831. https://journaleet.in/index.php/jeet/article/view/2508

Sivapragasam, C., & Natarajan, N. (2023). The Use of ICT at the Induction Level Towards Bringing Equity and Inclusion in HEIs of India. In Handbook of Research on Implementing Inclusive Educational Models and Technologies for Equity and Diversity, 69-88. https://doi.org/10.4018/979-8-3693-0453-2.ch004

Suhonen, S. (2025). AI in Measurement-Based Learning: A Challenge for Assessment, an Opportunity for Tutoring. In 2025 6th International Conference of the Portuguese Society for Engineering Education (CISPEE), IEEE, 1-5. https://doi.org/10.1109/CISPEE64787.2025.11124148

Sundar, M. S., Natarajan, N., & Vasudevan, M. (2020). A handy tool for forecasting population to aid estimation of water demand. Indian Journal of Geo-Marine Sciences, 49(09), 1587-1592. http://nopr.niscair.res.in/handle/123456789/55517

Xiao, X., Li, Y., He, X., Fang, J., Yan, Z., & Xie, C. (2025). An assessment framework of higher-order thinking skills based on fine-tuned large language models. Expert Systems with Applications, 272, 126531. https://doi.org/10.1016/j.eswa.2025.126531

Zheng, S., Huang, J., & Chang, K. C. C. (2023). Why does chatgpt fall short in providing truthful answers?. arXiv preprint, arXiv:2304.10513. https://arxiv.org/abs/2304.10513

Zhai, X., & Nehm, R. H. (2023). AI and formative assessment: The train has left the station. Journal of Research in Science Teaching, 60(6), 1390-1398. https://doi.org/10.1002/tea.21885

TABLE I
META DATA OF THE FORMATIVE ASSESSMENTS CONDUCTED AS CLOSED AND OPEN-BOOK TESTS

Particular	Feedback survey
Total number of questions	20
Number of questions on general background	3
Number of questions on preparation	2
Number of questions on performance	5
Number of questions on reflections	10

TABLE II
META DATA OF THE ONLINE SURVEY CONDUCTED FOR TEST FEEDBACK

Topic	RBT Level	Description of questions	Closed-book test		Open-book test	
			Number of questions	Marks assigned (%)	Number of questions	Marks assigned (%)
Fluid Properties	-	-	4	35	2	15
Fluid Pressure	-	-	2	25	3	20
Fluid Flow	-	-	4	40	5	65
-	R/F	-	2	15	2	15
-	U/C	-	4	25	4	25
-	Ap/C & Ap/P	-	3	40	3	40
-	An/C	-	1	20	1	20
-	Scenario-based		4	60	4	60
-	Image-based		1	10	2	20
-	Total number		10	-	10	-
-	Total marks		20	-	20	-

TABLE III
PERFORMANCE OF STUDENTS UNDER CLOSED-BOOK AND OPEN-BOOK TESTS WITH DIFFERENT RBT LEVELS OF QUESTIONS

RBT level	Closed-book test			Open-book test		
	Total marks assigned (%)	Average number of students attempted (%)	Average marks obtained (%)	Total marks assigned (%)	Average number of students attempted (%)	Marks obtained (%)
R/F	15	81.25	4.17	15	100	91.67
U/C	25	100	55.00	25	100	97.50
Ap/C	20	100	28.13	15	100	50.00
Ap/P	20	62.5	6.25	25	100	70.00
An/C	20	100	25.00	20	100	78.13

TABLE IV
DESCRIPTIVE STATISTICAL COMPARISON OF STUDENTS' PERFORMANCE (% MARKS) UNDER CLOSED-BOOK AND OPEN-BOOK TESTS WITH DIFFERENT RBT LEVELS OF QUESTIONS

RBT level	Closed-book test			Open-book test		
	Median	Standard deviation	Variance	Median	Standard deviation	Variance
R/F	0	11.79	138.89	100.00	15.43	238.10
U/C	60	17.73	314.29	100.00	7.07	50.00
Ap/C	25	20.86	435.27	66.67	35.63	1269.84
Ap/P	0	11.57	133.93	80.00	28.28	800.00
An/C	25	13.36	178.57	75.00	20.86	435.27

TABLE V
COMPARISON OF PAIRED T-TEST RESULTS FOR THE STUDENTS' PERFORMANCE UNDER CLOSED-BOOK AND OPEN-BOOK TESTS

Particular	Closed-book test	Open-book test
Degree of freedom	4	4
t-stat	10.48	2.68
P(T<=t) one-tail	2.3E-4	0.03
t-critical one-tail	2.13	2.13
P(T<=t) two-tail	4.7E-4	0.06
t-critical two-tail	2.78	2.78

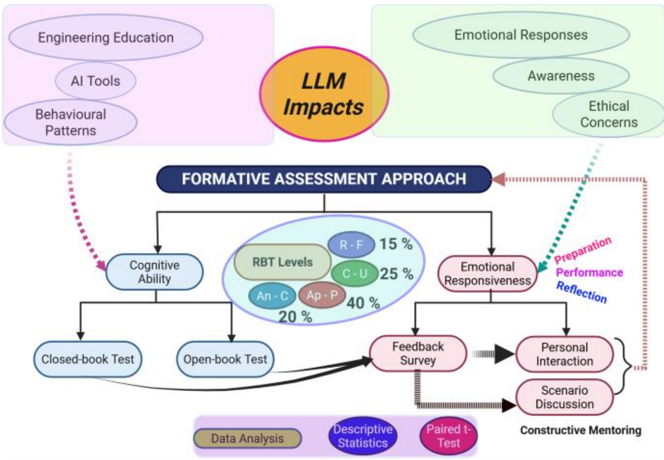


Fig. 1. Methodology for conducting formative assessment for assessing the LLM impacts

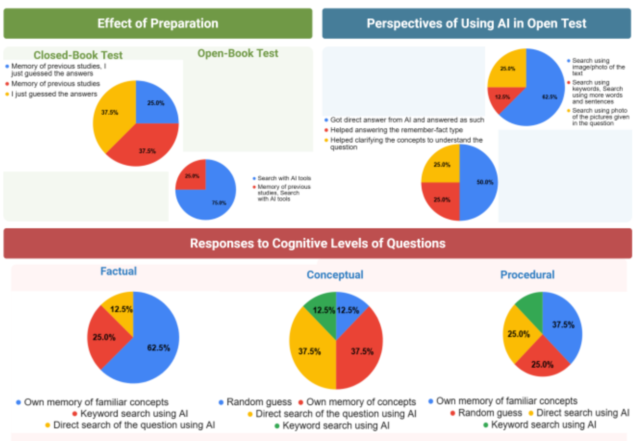


Fig. 2. Comparison of students' responses to the effect of preparation, perspectives on AI usage and to the cognitive levels of questions

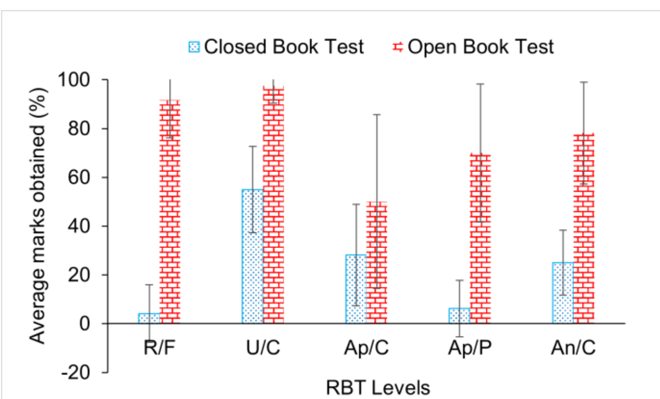


Fig. 3. Comparison of the average performance of the students during the closed-book and open-book tests

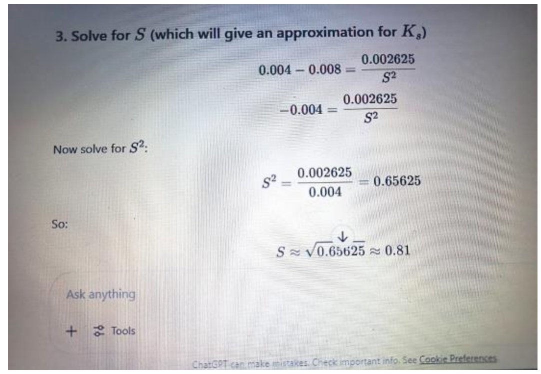


Fig. 4. A screenshot showing the evidence of an erroneous calculation made by ChatGPT

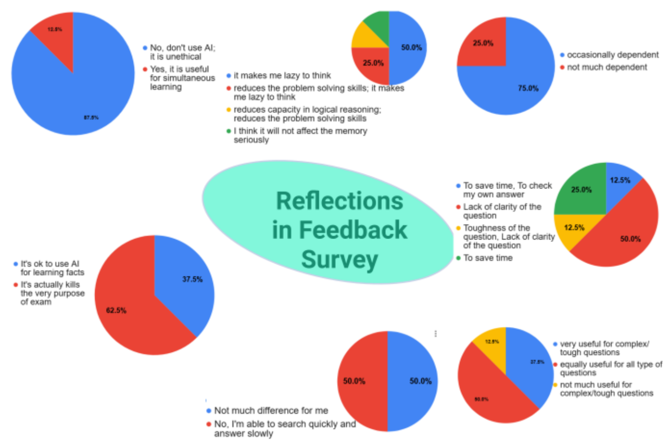


Fig. 5. Summary of reflections of the usage of AI tools in open-book tests based on the feedback survey results

SUPPLEMENTARY DATA

TABLE I
. LIST OF QUESTIONS USED IN THE CLOSED BOOK FORMATIVE ASSESSMENT

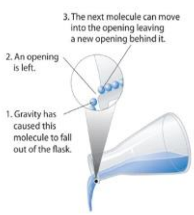
No.	Question	Options	Topic	RBT	Marks	Answer key
1	The dimension for kinematic viscosity is ___	$LM^{-1}T^{-1;}$ LT^{-2}; L^2T^{-1}; MLT^{-1}	Fluid properties	R/F	1	C
2.	Shear stress develops on a fluid element if the fluid	is at rest; subjected to linear acceleration; is inviscid, is viscous; and flow is non-uniform	Fluid properties	U/C	1	D
3.	If fluid is at rest in a container with a narrow mouth at a certain column height and the same fluid is at rest at the same column height in a container having a broad mouth, will the pressure be different at a certain depth from the fluid surface?	Pressure will be same for both; Pressure will be more for narrower mouth; Pressure will be less for narrower mouth; None of the mentioned	Pressure	Ap/C	1	A
4.	In which type of flow does the fluid velocity vary with time at a fixed point in the pipe?	Laminar; turbulent; uniform; steady	Flow	U/C	1	B
5.	Infer the correctness of the following statements. I. The velocity distribution is reverse of shear stress distribution in a pipe flow. II. Viscous flow may be mostly laminar in nature. III. Average velocity is half on the maximum velocity in pipe flow. IV. Change in pressure is called head loss and is due to the frictional resistance between fluid layers	I, II, III and IV; I and III; II and IV; I and IV	Flow	U/C	1	A
6.	Express the mass density, specific gravity, surface tension, and viscosity of water at standard temperature with proper units.		Fluid properties	R/F	2	997 kg/m^3; 1.0; 0.072 N/m; 0.0009 Pas
7.	Infer the reasons for (i) change in velocity profile during flow through a pipe of varying diameter; (ii) shape of an object affects the pressure distribution around it when immersed in a moving fluid.		Fluid flow	U/C	2	Principle of continuity; Resistance offered to the streamlines
8.	Fig.1 shows that the molecules in a liquid are in constant motion. Indicate the type of force(s) acts on the conical flask when it is tilted so that the result is a net flow of liquid out of the container.  Fig. 1: Fluid Flow		Fluid properties	Ap/C	3	Shear force – viscous layers – velocity gradient – flow
9.	A tank contains oil of specific gravity 0.85. Calculate the pressure at a depth of 4 m from the free surface. If the tank has two openings, one is wide and the other is narrow. Will there be any difference in the pressure at a given depth from the fluid surface? Justify your answer.		Hydrostatic pressure	Ap/P	4	P=hρg = 33.354 kPa No. hydrostatic law
10.	The law of conservation of energy can be written in terms of Bernoulli's equation as follows: $P_1/\rho g + V_1^2/2g + z_1 = P_2/\rho g + v_2^2/2g + z_2$ If there is a pump between two ends of a pipe, the pressure at the outlet can be calculated directly using this equation. Yes /No. Justify your answer.		Fluid flow	An/C	4	No, the equation has to be modified with the additional pressure created by the pump.

TABLE II
LIST OF QUESTIONS USED IN THE OPEN BOOK FORMATIVE ASSESSMENT

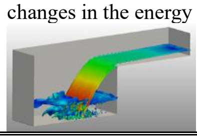
No.	Question	Options	Topic	RBT	Marks	Answer key
1	Which of the following cannot be the value of absolute pressure of a fluid at any point?	0 bar; -1 bar; 1.25 bar; 20 bar	Fluid properties	U/C	1	B
2.	What is the pressure in Pascals at a depth of 1 m below the water surface?	98100 Pa; 9810 Pa; 981 Pa; 1 Pa	Fluid pressure	Ap/P	1	B
3	Three beakers, 1, 2, and 3, of different shapes are kept on a horizontal table and filled with water up to a height h. If the pressure at the base of the beakers is P1, P2, and P3, respectively, which one of the following will be the relation connecting the three?	P1 > P2 > P3; P1 < P2 < P3; P1 = P2 = P3; P1 > P2 < P3	Fluid pressure	U/C	1	C
4.	Predict the causes major losses in pipes within the water supply network.	Changes in flow direction; Fittings and valves; Friction against the pipe's inner surface; Sudden variations in pipe diameter	Flow	R/F	1	C
5.	Identify the INCORRECT information about flow through pipes. I. Reynold's number (Re) is less than 2000 indicates parallel flow of streamlines. II. Reynold's experiment is conducted to find the transition between laminar flow and turbulent flow. III. The loss of pressure head is linearly proportional to the velocity when Re > 4000. IV. For higher Re, viscous action causes head loss due to friction	I; II; III; IV	Flow	U/C	1	C
6.	Define hydrostatic law in your words. Also, express it as an equation.		Hydrostatic pressure	R/F	2	Pressure changes with depth in a fluid at rest; P=hρg
7	Infer the reasons for (i) some insects walk on water; (ii) heavy ships float in ocean; (iii) steel plate sinks in water; (iv) air bubbles come up during boiling		Buoyancy and floatation	U/C	2	(i) surface tension of water film supports their weight; (ii) buoyant force counteracts gravity; (iii) more density than water; (iv) low density and high energy
8.	Select the type of flow for the conditions given below. (a)  (b)  (c) 		Fluid flow	Ap/C	3	Non-uniform and steady, pressurised flow; Non-uniform, unsteady and natural flow; Viscous, laminar, uniform, unsteady flow
9	A pipe of diameter 150 mm carries water at 0.25 m³/s. Calculate the velocity of flow. Also, illustrate the velocity profile inside the pipe for a given cross-section.		Fluid flow	Ap/P	4	Q=AV; v= 14.15 m/s
10.	A real-time picture of 3D flow indicates the distribution of velocity and acceleration in all directions of space and time, including rotation and translation is given below. Attribute the changes happening to the nature of flow in terms of Reynold's number. Also, compare the changes in the energy of the fluid before and after the fall. 		Fluid flow	An/C	2	laminar to turbulent; decrease in potential energy and an increase in kinetic energy

TABLE III
LIST OF QUESTIONS AND THE OPTIONS USED IN THE ONLINE FEEDBACK SURVEY

Question	Options
What helped you to answer the questions in the closed-book test?	Memory of previous studies; Immediate preparation; I just guessed the answers; Other:
What helped you to answer the questions in the open-book test?	Memory of previous studies; Immediate preparation; Google search without AI; Search with AI tools; Other:
How did you approach answering remember-facts, units, values - type of questions?	Own memory of familiar concepts; Random guess; Direct search of the question using AI; Keyword search using AI; Other:
How did you approach answering concepts-definition, scenario, application type of questions?	
How did you approach answering procedural - solving sums, steps, application type of questions?	
In which way AI helped you in answering the questions in the open-book test?	Helped answering the remember-fact type; Helped clarifying the concepts to understand the question; Got direct answer from AI and answered as such; Got confusing answers and did not use the result as such; Other:
In which way did you use AI for the test?	Search using keywords; Search using more words and sentences; Search using image/photo of the text; Search using voice of the text; Search using photo of the pictures given in the question; Other:
In which way do you think the use of AI will affect your own thinking and processing capacities?	I think it will not affect the memory seriously; Frequent use of AI will reduce the brain capacity in logical reasoning; frequent use of AI will reduce the problem-solving skills; it makes me lazy to think even small things that are already known to me; AI is a gift for persons like me who can't remember anything for a long while; Other:
How much are you dependent on AI tools in your regular academic and non-academic studies?	Very much dependent; occasionally dependent; not much dependent; not at all dependent
In your opinion, what prompted you to use AI for answering the questions in the open-book test?	Toughness of the question; Lack of clarity of the question; To save time To check my own answer; Other:
For which type of questions was the use of AI tools found to be more useful?	very useful for simple questions; not much useful for simple questions; very useful for complex/ tough questions; not much useful for complex/tough questions; equally useful for all type of questions; not much useful for any type of questions
Do you think that using AI in the exam reduces your available time for writing the answers?	Yes, it's taking more time in searching; No, I'm able to search quickly and answer slowly; Not much difference for me; I didn't use any AI tools for the test
On an average, how much time did you take for answering the questions in the closed-book test?	10-15 min; 15-30 min; More than 30 min
On an average, how much time did you take for answering the questions in the open-book test?	
How do you feel about using AI tools in exams?	It's helpful for me to answer the questions; It's ok to take help from AI for remembering the facts; It's helpful for correcting my own answer; It's actually kills the very purpose of the exam; Other:
Do you recommend the use of AI for your internal examinations in the future?	Yes, I think it'll useful for improving my understanding while answering the questions; No, I think it's wasting time for answering even the simple questions; Yes, it's advisable to use AI for higher order questions to avoid bias in assessment; No, it's not advisable to use AI for any exams because it is unethical; Other:
What steps would you take to minimize the impacts of AI in your own cognitive development?	[descriptive type]