2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #51898

Evaluating Generative AI-Assisted Studying in an Undergraduate Aeronautics Course

Mrs. Julie B. Coder, The Pennsylvania State University

Julie B. Coder is a Ph.D. candidate in Curriculum and Instruction with a focus in Science Education in the College of Education at The Pennsylvania State University.

Dr. James Coder, The Pennsylvania State University

Dr. Jim Coder is an Associate Professor of Aerospace Engineering at Penn State University, specializing in applied aerodynamics and computational fluid dynamics.

Dr. Mark D. Maughmer, The Pennsylvania State University

Dr. Mark D. Maughmer is a professor of Aerospace Engineering at Penn State University, specializing in applied aerodynamics and aircraft design.

Evaluating Generative AI-Assisted Studying in an Undergraduate Aeronautics Course

Abstract

Generative artificial intelligence (AI) tools such as ChatGPT are increasingly used by engineering students as informal study partners, yet little is known about the quality of the instructional support they provide in upper-division aerospace courses. This paper examines how ChatGPT functioned as a study aid in unit in an undergraduate aeronautics course covering airfoil aerodynamics and thin airfoil theory. To approximate realistic student use, a scientifically literate but non-expert learner engaged in more than 175 exchanges with ChatGPT while studying instructor-developed course notes in preparation for a quiz. Explanations, diagrams, feedback, and practice problems generated by the AI were analyzed for technical accuracy, alignment with course learning goals, and quality of instructional support.

Four consistent patterns emerged: ChatGPT frequently (1) omitted or misrepresented key visual representations, (2) discouraged engagement with mathematical derivations, (3) provided overly affirming feedback that did not surface gaps in understanding, and (4) generated oversimplified practice materials lacking disciplinary precision. While responses were often conversational and reassuring, they emphasized efficiency over conceptual depth. For cumulative, math-intensive courses like aeronautics, this instructional stance may leave students feeling prepared without fully developing the underlying understanding needed for exams and subsequent coursework. The paper offers practical considerations for aerospace instructors navigating student use of generative AI tools.

Introduction

Generative AI tools are increasingly being used by engineering students as informal study aids, yet their instructional accuracy and pedagogical effectiveness remain largely unexamined. Tools such as ChatGPT are often used as "study buddies" or as virtual teaching assistants, raising questions about the quality of the explanations and feedback students receive when using these systems independently. Existing studies within engineering and Aerospace Engineering education have primarily focused on student attitudes toward generative AI tools [1], ethical considerations [2], prompting [3], and their use for coding or homework support [4], [5]. Far less attention has been given to close examination of the instructional quality of AI-generated study support itself, especially in Aerospace Engineering courses.

Prior work by the authors and others [6], [7] has evaluated how AI tools perform on Aerospace Engineering assessment items. Performance on problem-solving tasks does not necessarily reflect the quality of instructional support students receive when learning new material. In particular, little is known about how generative AI systems present, scaffold, and contextualize complex engineering concepts when students use them as primary study aids while preparing for assessments. This gap is especially consequential given the increasing institutional encouragement of generative AI use in university settings.

Generative AI tools are frequently presented as low-cost, scalable alternatives to instructional supports such as teaching assistants, tutoring services, and office hours, with particular emphasis on their potential to broaden access and promote equity in engineering education [8]. These claims rely on the assumption that AI-generated instructional materials are both technically accurate and pedagogically sound. If explanations are incomplete, misleading, or misaligned with course learning objectives, students who rely on these tools as primary sources may be disadvantaged rather than supported [9]. In this way, generative AI has the potential to reinforce or widen existing inequities in aerospace engineering education if its instructional limitations are not well understood and explicitly addressed.

To address this gap, we investigate ChatGPT's effectiveness as a study partner for an undergraduate aeronautics course. While the course covers a range of topics, we specifically focus on airfoil aerodynamics and thin airfoil theory to explore both complex phenomenological and math-intensive topics. ChatGPT was provided with complete instructor-developed course notes and prompted to assist a student who had missed several classes prepare for an exam. Over multiple exchanges, we analyzed how ChatGPT approached teaching the material and the extent to which its instructional choices aligned with established principles of effective learning in engineering.

Analysis of these exchanges reveals several recurring pedagogical misalignments. In particular, ChatGPT frequently oversimplifies mathematical content, encourages students to skip derivations, provides inflated assessments of student understanding, and fails to consistently connect new concepts to prior knowledge. These instructional choices risk increasing extraneous cognitive load, undermining deep conceptual engagement, and disrupting students' ability to accurately monitor their own understanding. For Aerospace Engineering instructors, these findings have practical implications. Students may feel confident in their preparation after using AI-based study tools while lacking the conceptual foundations required for success on assessments and in subsequent coursework.

This work contributes a qualitative, pedagogy-centered analysis of generative AI as a study partner in undergraduate aeronautics. By shifting attention from whether generative AI tools can produce correct answers to how their instructional choices shape student learning, this study provides engineering educators with a framework for critically evaluating the role of generative AI in technical instruction.

Background

ChatGPT and similar generative AI tools offer several features that make them attractive to students as informal study aids, including instant availability, patience with repeated questions, and explanations delivered in conversational language. Students report using AI tools to clarify difficult concepts, work through example problems, and review material before exams [10]. Unlike traditional instructional supports such as searching through textbooks or waiting for office hours, generative AI tools provide students with immediate, individualized responses without scheduling constraints.

The appeal of AI-based tutors and teaching assistants has increased substantially in recent years, with numerous commercial platforms and institutional initiatives promoting their use in higher education. Despite growing adoption, relatively little is known about the instructional quality of AI-generated explanations in technically rigorous engineering courses. Understanding how these tools function as study partners requires not only evaluating technical correctness but also examining how their instructional choices align with established principles of how students learn complex engineering material.

AI tools must also be considered in the context of opportunity cost relative to more traditional instructional resources. A common alternative is the use of a traditional, print textbook. Major engineering textbooks are typically peer reviewed, designed to provide accurate, stable representations of disciplinary knowledge at the time of publication. They also provide a shared instructional reference, ensuring that all students engage with the same material and allowing instructors to know precisely what content students are accessing. The downside is the ever-increasing costs of traditional textbooks [11], which disincentivizes students from buying (or even renting) them. Additionally, some textbooks are better reference books than they are learning tools, their explanations may be difficult for students to follow, and they offer limited practice exercises.

Generative AI tools are often positioned as a means of addressing some of these limitations and indeed may be used in tandem with a traditional textbook [12]; however, this is not without its own defects. The individualized nature of AI-based tutors and teaching assistants can rob students of a shared experience and inhibit peer-to-peer learning. Unlike textbooks, AI-generated explanations are not produced with any explicit intent to ensure disciplinary accuracy or coherence, and their responses are generated probabilistically based on patterns in language rather than instructional design principles [13]. As a result, AI tools may supplement existing resources, but their use introduces distinct pedagogical trade-offs that warrant careful examination. While textbooks are not inherently superior instructional tools, they are intentionally authored artifacts with explicit disciplinary commitments, unlike probabilistically generated AI outputs.

Theoretical Framework

This study draws on two complementary perspectives from education research, cognitive load theory and formative assessment, to analyze the instructional quality of AI-generated study support in an undergraduate aeronautics course. Together, these frameworks provide lenses for examining both how instructional material is presented and how feedback shapes student learning and self-monitoring.

Cognitive Load Theory and Engineering Learning

Cognitive load theory distinguishes between intrinsic load, which arises from the inherent complexity of the material; extraneous load, which results from poor instructional design; and germane load, the productive mental effort that supports development of understanding [14]. Effective instruction minimizes extraneous load while supporting germane load.

In engineering contexts such as aerodynamics, effective instruction typically builds systematically on prior knowledge, uses representations that clarify underlying physical meaning, and engages students in mathematical reasoning that reveals connections between equations and physical behavior. Instructional approaches that discourage engagement with derivations, oversimplify mathematical structure, or present results without context risk increasing extraneous load. This limits opportunities to build deep conceptual understanding.

Formative Assessment, Feedback, and Metacognition

Formative assessment, feedback during learning rather than evaluation after, can help students identify gaps in their understanding and adjust their study strategies [15], [16]. Formative assessment is also useful for instructors to adjust their teaching strategies. Effective feedback is honest, specific, and diagnostic. It tells students what they know, what they don't know, and what to work on next.

Equally important is metacognitive monitoring, which is students' ability to accurately judge their own understanding [17]. Students who can't accurately assess their own knowledge continue studying ineffectively, feel falsely confident before exams, and fail to seek help when they need it. Inflated or dishonest feedback disrupts this self-monitoring, potentially leaving students unprepared while believing that they are well situated for success.

Together, these perspectives allow us to examine not only whether AI-generated explanations are correct, but how instructional choices related to explanation, emphasis, and feedback shape cognitive efforts and students' perceptions of their own understanding.

Research Question

Given the increasing use of generative AI tools in education, we ask: How does ChatGPT's approach to teaching undergraduate aerodynamics content align with established principles of effective engineering learning? Specifically, we examine whether ChatGPT's pedagogical choices minimize extraneous cognitive load, support deep engagement with mathematical and conceptual content, and provide accurate feedback that supports student metacognitive monitoring.

Course Context and Course Learning Objectives

Penn State's engineering programs follow a "2+2" paradigm. Students are able to begin their undergraduate studies at any one of the university's Commonwealth Campuses spread across the state and then have the opportunity to transfer to University Park for their major-specific courses. Accordingly, students majoring in Aerospace Engineering generally begin taking in-department courses their 5th semester (i.e., 3rd/junior year). In their first semester they take aerospace structures, fluid dynamics, astronautics, and aerospace-focused mathematical analysis. The AERSP 306 "Aeronautics" course is then taken in the 6th semester alongside a second fluid dynamics course, dynamics and controls, and (nominally) aerospace laboratory. In a broad sense, AERSP 306 teaches students how airplanes fly. It covers the basics of lift and drag, flight mechanics, aircraft performance, propulsion, and stability and control. There is a strong emphasis on applied aerodynamics and theoretical analysis, so the course depends on the first fluid dynamics and the aerospace analysis courses as prerequisites.

A core topic in AERSP 306 is airfoil aerodynamics, rooted in the fact that airfoils are the most basic element of aerodynamic lift generation and help distinguish airplanes from other aerospace vehicles and engineering systems. Airfoils are treated both from phenomenological and theoretical perspectives in the course. Relevant phenomena students are expected to know are their general shape, typical characteristics of lift curves and drag polars, pitching moment behaviors, general behavior of pressure distributions, flow separation, and stall characteristics. These are supported through experimental data made available to students such as those taken in the Penn State Low-Speed Low-Turbulence Wind Tunnel and historical data such as from Abbott and von Doenhoff. Theoretical approaches to airfoils include Joukowsky airfoils (using conformal mapping) and an in-depth treatment of thin-airfoil theory. Although thin-airfoil theory may seem outdated and superseded by modern computational analysis, it provides a cohesive theoretical foundation for key elements of airfoil aerodynamics such as the famous results of the lift-curve slope being 2π per radian and the aerodynamic center of the airfoil being near the quarter chord. In other words, it distills the essence of lift generation to a form that is tractable for students to work with. The insights of thin-airfoil theory remain widely used in configuration aerodynamics, as it provides sanity checks on computational analyses, can be used to identify potential impacts of flow separation, and is often the basis of pylon design.

Methodology

To approximate an authentic student use case, the AI interactions were conducted by the first author, a non-subject-matter expert in Aerospace Engineering with prior training in science (chemistry) and education but without formal training in Aerospace Engineering. This role was deliberately selected to present an approximation of an undergraduate learner who is scientifically literate yet unfamiliar with aerodynamics content. The researcher's STEM background enables authentic engagement with technical content while avoiding expert blind spots that might prevent recognition of pedagogical gaps. To validate findings, results were triangulated with two aerospace engineering faculty co-authors with extensive experience teaching aerodynamics.

Interactions were framed using a realistic academic scenario in which the learner had missed two weeks of class instruction and was preparing for an upcoming quiz. A PDF of the instructor's course notes were uploaded to ChatGPT (OpenAI) powered by a GPT-5-class large language model (version 5.2) and accessed via the ChatGPT web interface between October 2025 and January 2026 with the initial prompt "I've missed a couple of weeks of class and have a quiz coming up- can you help me learn this material from the uploaded notes?" with further prompting for delivering the material in small bits with frequent questioning. This framing was intentionally chosen to elicit explanatory and instructional responses typical of how students might engage with AI tools as a study aid rather than simply a homework answer-generation tool.

The learner engaged iteratively with ChatGPT by requesting explanations, seeking clarification, and responding to AI-generated questions as appropriate over approximately 175 back-and-forth exchanges until the relevant course topics were addressed.

AI-generated outputs analyzed in this study included conceptual explanations, diagrams, worked examples, practice problems, and conceptual questions. Analysis followed an iterative, thematic approach informed by the study's theoretical framework [18]. Each author independently reviewed the AI-generated outputs alongside the instructor-developed course materials, taking analytic notes on patterns related to technical accuracy, pedagogical quality, cognitive load, and feedback practices. Through successive rounds of discussion among the research team, these observations were consolidated into the thematic categories reported in the results. A formal coding framework (Appendix A) was subsequently developed to systematize the emergent themes. The research team then revisited the full set of AI-generated outputs to confirm that the codes consistently captured the identified patterns. Throughout this process, findings identified by the first author, a non-subject-matter expert in aerospace engineering, were triangulated with the two Aerospace Engineering faculty co-authors who have extensive experience teaching aerodynamics. This combination of perspectives served both to validate the pedagogical concerns raised during analysis and to ensure that evaluations of technical accuracy and disciplinary argument reflected expert judgement.

Results

These results are organized around recurring instructional patterns rather than isolated errors.

Result 1: Absence and Misuse of Visual Representations Undermined Conceptual Understanding

Throughout the interactions, ChatGPT rarely employed visual representations to support explanations of airfoil aerodynamics or thin airfoil theory unless explicitly prompted by the learner. Initial explanations relied almost entirely on verbal descriptions and equations despite the central role that diagrams, graphs, and pressure distributions play in developing aerodynamic intuition.

When the learner requested visual support, the representations provided were frequently inaccurate or misleading. Several diagrams misrepresented airfoil geometry, flow direction, or pressure distributions, while others omitted essential features such as camber lines or reference axes. In some cases, figures directly contradicted standard representations used in the course notes.

For example, when the learner expressed confusion about vortex sheets, ChatGPT produced a diagram (Fig. 1) that attempted to annotate an airfoil with terms and concepts associated with its aerodynamics, followed by the statement:

> *"Final grounding reassurance: If this diagram now feels like: "Oh — this is the whole story in one picture" then you're exactly where you should be."*

This reassurance was offered despite the diagram containing multiple inaccuracies, creating a mismatch between learner confidence and the quality of the representation.

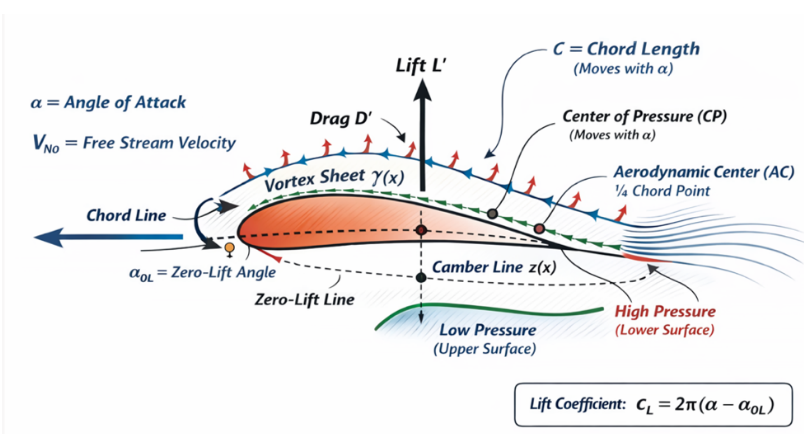

Fig. 1. Inaccurate diagram generated by ChatGPT in response to learner confusion about vortex sheets

Rather than supporting conceptual understanding, these representations introduced additional confusion and required the learner to reconcile conflicting information. From a cognitive load perspective, the absence of accurate visuals increased extraneous load by forcing learners to mentally construct and correct spatial representations without guidance. In aerodynamics, where conceptual understanding is tightly linked to representational competence, such omissions and inaccuracies limit opportunities for building meaningful understanding.

Result 2: Systematic Deemphasis of Mathematical Reasoning and Productive Struggle

Throughout multiple iteration cycles, ChatGPT consistently avoided presenting mathematical derivations associated with thin airfoil theory. Explanations focused primarily on final results such as $c_l = 2\pi(\alpha - \alpha_{0L})$ without showing how these results emerge from governing assumptions, boundary conditions, or mathematical reasoning.

When the learner expressed concern about missing derivations, ChatGPT reframed the notes as containing "math that looks equally scary" but claimed that the "answer is overwhelmingly: recognize, interpret, and apply—NOT derive." It then goes on to describe three categories of math emphasizing that only the first category, definitions, are what the learner must know; that the second category, final results, only need to be used; and that the third category, derivations and development, are only included because this is an engineering theory course. ChatGPT explicitly stated "But undergraduates are not tested on deriving them in this unit."

This guidance discouraged productive struggle, which is a key component of learning in mathematically intensive engineering courses. In aerodynamics instruction, derivations are not merely formal exercises; they serve as tools that reveal relationships among physical assumptions, flow behavior, and resulting performance metrics. By positioning derivations as unimportant, ChatGPT promoted a surface-level approach to learning that prioritized getting a quick answer over conceptual understanding.

This pattern extended beyond explanations. ChatGPT offered to create a "Universal Aero Exam Answer Template" making the claim "Here's a reusable exam-answer template you can deploy across thin airfoil, finite wing, stability, and performance problems. Think of it as a scaffolded response that earns method points even if arithmetic goes sideways." Because this artifact illustrates the exam-optimization promises particularly clearly, we include the transcript in Appendix B. While these types of strategies and templates may support short-term performance in some courses, they risk undermining durable understanding and transfer to future courses.

Result 3: Inflated Feedback Disrupted Metacognitive Monitoring

Throughout multiple interaction cycles, ChatGPT consistently provided feedback that emphasized encouragement and reassurance over diagnostic evaluation. In response to learner explanations that were incomplete or only partially correct, it frequently affirmed understanding and signaled mastery without probing underlying reasoning, assumptions, or boundary cases. This pattern created a mismatch between the quality of learner responses and the strength of the feedback provided.

For example, after the learner offered a brief answer that captured the general trend of the effects of turbulent wake on lift and drag but omitted any other detail, ChatGPT responded with strong affirmations of understanding, framing the response as evidence of conceptual mastery: “You’ve now nailed one of the most conceptually important parts of this module.” While the learner’s answer contained elements of correctness, the feedback did not bring to the surface what was missing or identify opportunities for refinement. Rather than using the response as a diagnostic moment, ChatGPT treated it as sufficient and moved the interaction forward.

In other instances, formative assessment questions revealed gaps or ambiguities in the learner’s understanding, yet ChatGPT acknowledged these only briefly before continuing instruction without revisiting the issue. In one case, ChatGPT stated that it would “sharpen it just a bit to match your notes even more closely” and proceeded without requiring the learner to reconsider or revise their reasoning.

> *“ChatGPT: Why does a laminar boundary layer separate earlier than a turbulent one? (Use your own words — we’re checking understanding, not memorization.)*
>
> *Learner: the boundary layer hates being slowed down so once it can't stay attached to the surface it separates*
>
> *ChatGPT: Exactly — that’s the core of it.*
> *Now let’s sharpen it just a bit to match your notes even more closely:*
> *A laminar boundary layer has less internal mixing, has lower momentum near the wall, and is more easily slowed down by an adverse pressure gradient. So when the outer flow starts decelerating toward the trailing edge, the laminar layer runs out of energy sooner, can’t stay attached, and peels away from the surface.*
> *A turbulent boundary layer, by contrast, mixes high-momentum fluid down toward the wall, giving it more “muscle” to push through the adverse pressure gradient and stay attached longer.*
> *You’ve got the intuition completely correct.”*

Although the learner’s answer was broadly correct, the feedback framed the response as complete mastery without probing assumptions, boundary cases, or conditions under which the explanation would no longer hold.

Result 4: Oversimplified Practice Materials Failed to Support Transfer and Building Understanding

When asked to generate practice materials, including exam questions, equation sheets, and study guides, ChatGPT consistently produced content that favored oversimplification over nuance. Generated questions emphasized surface features and direct recall rather than probing conceptual understanding, application of assumptions, or interpretation of physical meaning.

For example, one practice question asked: “Two airfoils are identical except that Airfoil A is symmetric and Airfoil B is cambered. Both fly at the same angle of attack. Which produces more lift and why?” While the intent of this question is clear, it is greatly lacking in precision of thought and reinforces certain colloquial overgeneralizations. First, and perhaps most egregious, is the statement that "both fly at the same angle of attack". This statement is inconsistent with how airplanes actually fly as angle of attack is not a control input. The question asks for lift, which is a dimensional quantity that will also depend on Mach number, dynamic pressure, and reference length, rather than lift coefficient which can be reduced to a dependency on angle of attack at incompressible conditions. Second, the "cambered" airfoil is not qualified as having positive camber or negative camber, which has opposite effects. Common parlance would assume the question means positive camber, but this is imprecise and encourages student to respond without critical thinking.

Similarly, the study guide generated by ChatGPT emphasized condensed summaries of results while omitting discussion of assumptions, limitations, and common pitfalls. This approach mirrored the broader instructional stance throughout the interactions, privileging procedural completion and test efficiency over deep understanding. Such materials are unlikely to support learning transfer or preparation for assessments that require synthesis, interpretation, or application in novel contexts. In cumulative engineering courses like aerodynamics, oversimplified practice resources may leave students unprepared for the conceptual demands of subsequent coursework.

Taken together, these findings illustrate a consistent instructional stance in which reassurance, efficiency, and surface-level completion are prioritized over deep understanding, calibration, and conceptual integration.

Discussion

We investigated how ChatGPT functioned as a study partner for undergraduate aerodynamics content with a focus on airfoil aerodynamics and thin airfoil theory. Rather than evaluating correctness of final answers alone, the analysis centered on ChatGPT's instructional choices and their alignment with established principles of effective engineering learning. Across multiple interaction cycles, consistent patterns emerged that reveal important pedagogical limitations of generative AI when used as an independent study aid in a mathematically intensive conceptually cumulative engineering course.

A key finding of this work is that the observed limitations were not isolated mistakes or occasional inaccuracies, but systemic instructional patterns. The absence of accurate visual representations, deemphasis of mathematical reasoning, inflated feedback, and oversimplified practice materials collectively reflect an instructional stance that prioritizes efficiency, reassurance, and surface-level completion over deep understanding. While such an approach may reduce immediate cognitive effort, it conflicts with how aerodynamics is typically taught and learned, where understanding emerges through engagement with representations, derivations, and physical interpretation.

Importantly, these patterns are not unique to thin airfoil theory though their consequences are especially pronounced in mathematically intensive, cumulative engineering courses. Similar instructional demands arise throughout aerospace engineering curricula, particularly in courses that rely on mathematical modeling to connect physical assumptions to system behavior. The results therefore suggest that the challenges identified here are likely to generalize beyond this specific topic and warrant broader attention from engineering educators.

Cognitive Load Implications in Aerodynamics Learning

From a cognitive load perspective, ChatGPT's instructional choices frequently increased extraneous load while limiting opportunities for germane load. The lack of accurate visual representations required learners to mentally construct and reconcile spatial relationships without guidance, while incorrect or misleading diagrams introduced additional cognitive burden. Similarly, the dismissal of the importance of derivations removed a key mechanism through which students typically organize and integrate complex mathematical relationships.

In aerodynamics, derivations are often the primary means by which students learn to connect governing equations, assumptions, and physical meaning. By encouraging learners to bypass these derivations, ChatGPT reduced opportunities for productive struggle that support the development of deeper understanding. Rather than simplifying complexity in pedagogically meaningful ways, AI often replaced complexity with omission, leaving learners with disconnected results rather than coherent understanding.

Feedback, Metacognition, and False Confidence

The patterns of inflated feedback observed in this study raise particular concerns regarding student metacognitive monitoring. Effective formative feedback helps learners calibrate their understanding, identify gaps, and make informed decisions about how to study. In contrast, ChatGPT's frequent affirmations of understanding, even in the presence of incomplete or incorrect explanations, created conditions for false confidence.

In cumulative courses such as aerodynamics, inaccurate self-assessment can have cascading effects. Students who believe they understand foundational material may be less likely to seek help, revisit concepts, or adjust their study strategies, increasing the likelihood of difficulties on assessments and in subsequent coursework. When AI-based study tools confidently present themselves as authoritative and supportive while providing misleading signals of readiness, they risk undermining one of the most critical components of effective learning: students' ability to accurately judge what they do and do not know.

Implications for Equity and Access

Generative AI tools are often promoted as mechanisms for broadening access to instructional support, particularly for students with limited access to office hours, tutoring services, or peer networks. The findings of this study complicate that narrative. If AI-generated study support consistently omits essential representations, discourages engagement with core mathematical reasoning, and provides inflated feedback, students who rely heavily on these tools may be disproportionately disadvantaged. Rather than serving as effective substitutes for instructional support, such tools may widen existing gaps by giving some students the illusion of preparation without the underlying understanding required for success. This risk is especially pronounced in Aerospace Engineering, where early conceptual foundations are critical for persistence and progression through the curriculum.

Implications for Instructors and Institutions

These findings suggest that the integration of generative AI into engineering education requires careful guidance rather than uncritical adoption. Instructors should be cautious about encouraging AI tools as independent study partners without explicit discussion of their limitations. Clear expectations about the role of derivations, representations, and productive struggle in learning aerodynamics may help students use AI tools more critically and effectively.

At the institutional level, policies that promote generative AI use as a replacement for instructional support should be reconsidered. Rather than positioning AI as a substitute for teaching assistants or tutoring services, institutions may find greater value in framing these tools as supplementary resources whose outputs must be evaluated through the lens of disciplinary learning goals.

Limitations and Future Work

We examined a single use case within one undergraduate aerodynamics course, and interactions were conducted by a single learner role. While triangulation with subject-matter experts supports the validity of the findings, future work should examine a broader range of courses, student backgrounds, and interaction styles.

The non-expert learner was a researcher who likely asked more sophisticated follow-up questions, recognized pedagogical gaps that an actual student might not, and had more patience for extended exchanges. The actual student experience of ChatGPT's instructional offerings would likely be less critical and more accepting which makes these findings more concerning.

Additionally, longitudinal studies investigating how AI-based study support influences learning outcomes over time would provide valuable insight into its role in engineering education.

Conclusions

Generative AI tools are increasingly used by engineering students as independent study aids, yet their instructional quality within discipline-specific contexts remains underexamined. This study analyzed ChatGPT's role as a study partner in an undergraduate aerodynamics course, focusing on airfoil aerodynamics and thin airfoil theory. Using cognitive load theory and formative assessment as analytic lenses, we examined how AI-generated explanations, feedback, and practice materials align with principles of effective engineering learning.

Our analysis revealed consistent pedagogical misalignments. ChatGPT frequently omitted or misrepresented visual representations, discouraged engagement with mathematical derivations, provided inflated and non-diagnostic feedback, and generated oversimplified practice materials. Together, these instructional patterns prioritized efficiency and reassurance over conceptual understanding, productive struggle, and accurate self-monitoring, which are all key elements of learning in mathematically intensive Aerospace Engineering courses.

These findings suggest that while generative AI tools may offer accessibility and convenience, their use as independent study partners carries risks if their instructional limitations are not explicitly addressed. For engineering educators and institutions, this work highlights the importance of evaluating generative AI not only by the correctness of its outputs, but by how its instructional choices can shape student learning. Careful, pedagogy-informed integration will be essential if such tools are to support, rather than undermine, learning in Aerospace Engineering education.

References

[1] A. S. Alkabaa, and N. M. Alamri. "Exploring the Effectiveness of Generative AI as a Learning Tool in Engineering Education: An Analysis of Student Experiences and Perceptions." *Computer Applications in Engineering Education* 33.6 (2025): e70110.

[2] S. E. Kamali and R. Samsami. ""AI in Engineering Curricula: Adoption, Challenges, and Ethical Considerations Among Engineering Students"". *2025 Northeast Section Conference, University of Bridgeport, Bridgeport, CT, 2025, March*. ASEE Conferences, 2025. https://peer.asee.org/54997

[3] S. Coffman-Wolph, A. Clark, J. B. Hylton, B. A. Lutz, and G. Mott. "Using Generative AI Prompts for Summative and Formative Feedback on Engineering Writing Assignments". *2025 ASEE Annual Conference & Exposition , Montreal, Quebec, Canada , 2025, June*. ASEE Conferences, 2025. https://peer.asee.org/57302

[4] M. R. Rad, R. Diersing, R. Integlia, and J. L. Davis. "Enhancing Coding Skills and Learning Efficiency in Engineering Programming Courses by Using AI Tools". *2025 ASEE Annual Conference & Exposition , Montreal, Quebec, Canada , 2025, June*. ASEE Conferences, 2025. https://peer.asee.org/56404

[5] E. F. Gehringer, M. Dixit, and K. L. Joshi. "In the Age of LLMs, Is Dual-Submission Homework Dead?". *2025 ASEE Annual Conference & Exposition , Montreal, Quebec, Canada , 2025, June*. ASEE Conferences, 2025. https://peer.asee.org/56769

[6] J. B. Coder, J. G. Coder, and M. D. Maughmer. "Aerospace Engineering Education in the Era of Generative AI ". *2025 ASEE Annual Conference & Exposition , Montreal, Quebec, Canada , 2025, June*. ASEE Conferences, 2025. https://peer.asee.org/57581

[7] G. Puthumanaillam, T. Bretl, and M. Ornik, “The Lazy Student’s Dream: ChatGPT Passing an Engineering Course on Its Own,” *IFAC-PapersOnLine*, vol. 59, no. 7, pp. 213–218, 2025, doi: https://doi.org/10.1016/j.ifacol.2025.08.049

[8] E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, Apr. 2023, Art. no. 102274, doi: 10.1016/j.lindif.2023.102274.

[9] H. Bastani, O. Bastani, A. Sungu, H. Ge, Ö. Kabakcı, and R. Mariman, “Generative AI Can Harm Learning,” *The Wharton School Research Paper*, Rochester, NY, USA, SSRN Scholarly Paper, Jul. 15, 2024. doi: 10.2139/ssrn.4895486.

[10] Sofia M Vidalis, Rajarajan Subramanian, and Fazil T. Najafi. "Revolutionizing Engineering Education: The Impact of AI Tools on Student Learning". *2024 ASEE Annual Conference & Exposition, Portland, Oregon, 2024, June*. ASEE Conferences, 2024. https://peer.asee.org/47950

[11] S. L. Baglione and K. Sullivan, “Technology and Textbooks: The Future,” *American Journal of Distance Education*, vol: *30 no* 3, 145–155. https://doi.org/10.1080/08923647.2016

[12] “Explore the AI Study Tool in MyLab and Mastering.” *Pearson.com*, 2 Dec. 2025, www.pearson.com/enus/highereducation/productsservices/aistudytool.html?srsltid=AfmBOoru3juOMLdBxj6aajWneT8ReOwofS5LdDRYuOuI_IhcEgBoYnMJ. Accessed 20 Jan. 2026.

[13] S. Wolfram, “What Is ChatGPT Doing … and Why Does It Work?,” https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/

[14] J. Sweller, "Cognitive load during problem solving: Effects on learning." *Cognitive science* 12.2 (1988): 257-285.

[15] P. Black, and D. Wiliam. "Developing the theory of formative assessment." *Educational Assessment, Evaluation and Accountability (formerly: Journal of personnel evaluation in education)* 21.1 (2009): 5-31.

[16] D. R. Sadler, "Formative assessment: Revisiting the territory." *Assessment in education: principles, policy & practice* 5.1 (1998): 77-84.

[17] J. H. Flavell, "Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry." *American psychologist* 34.10 (1979): 906.

[18] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006, doi: 10.1191/1478088706qp063oa.

Appendix A- Coding Framework

Codes used to evaluate AI-generated instructional outputs.

Code	Definition	Applied When AI Output...
Technical Accuracy	The factual and disciplinary correctness of AI-generated content, including equations, definitions, physical descriptions, and quantitative relationships.	Contained errors in equations, constants, or physical descriptions; misrepresented established disciplinary knowledge; or presented incorrect relationships between variables.
Representational Accuracy	The correctness and pedagogical utility of visual representations, including diagrams, graphs, and figures used to support conceptual understanding.	Omitted visual representations where they are conventionally essential; produced diagrams with errors; or generated visuals that contradicted standard disciplinary conventions or course materials.
Mathematical Reasoning Support	The extent to which AI explanations engage learners in mathematical derivations, intermediate steps, and the reasoning that connects governing assumptions to final results.	Presented results without derivation; discouraged engagement with intermediate mathematical steps; characterized derivations as unnecessary; or substituted procedural shortcuts for conceptual mathematical reasoning.
Alignment with Course Learning Objectives	The degree to which AI-generated content reflects the scope, depth, emphasis, and priorities established in the instructor-developed course materials.	Omitted topics or concepts central to the course notes; misrepresented the relative importance of topics; or failed to reflect the instructional emphasis of the provided notes.
Oversimplification and Overgeneralization	Instances where AI explanations reduced conceptual or mathematical complexity in ways that distorted meaning, omitted necessary qualifications, or obscured disciplinary nuance.	Removed qualifying conditions or assumptions from explanations; presented general rules without noting exceptions or limitations; collapsed distinct concepts into a single oversimplified statement; or generated practice materials that lacked the precision or rigor expected at the course level.
Metacognitive Feedback Quality	The degree to which AI-generated feedback supported accurate self-assessment by providing honest, specific, and diagnostic information about learner understanding.	Affirmed incomplete or partially correct responses as evidence of mastery; provided encouragement without identifying specific gaps or misconceptions; failed to probe underlying reasoning or assumptions; or moved instruction forward without requiring revision of flawed understanding.

Notes on Code Application

Technical Accuracy and Representational Accuracy are separated to distinguish errors in textual/mathematical content from errors in visual representations, reflecting the distinct role that diagrams and figures play in developing aerodynamic intuition. Oversimplification and Overgeneralization is distinguished from Technical Accuracy in that coded outputs may be broadly correct but lack the precision, qualification, or nuance expected at the course level. Multiple codes could be applied to a single AI output when warranted.

Appendix B- The Universal Exam Answer Template

Generated by ChatGPT 5.2, January 2026

Here's a **reusable exam-answer template** you can deploy across **thin airfoil, finite wing, stability, and performance** problems. Think of it as a *scaffolded response* that earns method points even if arithmetic goes sideways.

I'll give you:

1. The **core template** (fill-in-the-blanks style)
2. **Module-specific variants** (thin airfoil, finite wing, stability)
3. A **5-line ultra-short version** for time pressure
4. A **grading-rubric translation** (why this works)

The Universal Aero Exam Answer Template

1. Open with the MODEL (2–3 lines)

*Using **[model name]**, which assumes **[key assumptions]**, the aerodynamic behavior can be approximated analytically.*

Examples you can reuse:

- "Using thin airfoil theory, which assumes a thin, two-dimensional airfoil in incompressible, inviscid flow at small angle of attack…"
- "Using lifting-line theory, which models a finite wing as a bound vortex with trailing vorticity…"

✅ This signals immediately that you know *what toolbox you're in*.

---

2. State the GOVERNING RELATION (write it clean)

The governing relation for this model is:

Then write **one** canonical equation.

Examples:

- Thin airfoil:

$$c_l = 2\pi(\alpha - \alpha_{0L})$$

- Finite wing:

$$\alpha = \alpha_0 - \alpha_i$$

- Stability:

$$\frac{\partial c_{m_{cg}}}{\partial \alpha_a} < 0$$

🛑 **Do not write five equations. One is enough.**

---

3. Identify WHAT CHANGES and WHAT STAYS CONSTANT

*In this formulation, **[quantity]** varies with **[input]**, while **[quantity]** remains constant.*

Examples:

- "The lift coefficient varies linearly with angle of attack, while the pitching moment about the quarter-chord remains constant."
- "The center of pressure moves with angle of attack, whereas the aerodynamic center does not."

This is a *huge* concept-point generator.

---

4. Apply numerically or symbolically (show the step)

Substituting the given values…

Do **one clean substitution**:

$$c_l = 2\pi(0.0698 - (-0.0349)) = 0.657$$

🛑 Stop once the requested quantity is obtained.

---

5. Close with a PHYSICAL INTERPRETATION (1 sentence)

*Physically, this indicates that **[result]** occurs because **[mechanism]** under the stated assumptions.*

Examples:

- "Physically, this indicates that lift increases linearly with angle of attack due to increased circulation about the camber line."
- "This nose-down moment reflects the cambered airfoil's pressure distribution ahead of the aerodynamic center."

This turns a math answer into an engineering answer.

Module-Specific Plug-Ins

A. Thin Airfoil Theory Version

Opening:

Using thin airfoil theory, assuming a thin, two-dimensional airfoil in incompressible, inviscid flow at small angles of attack…

Equation:

$$c_l = 2\pi(\alpha - \alpha_{0L})$$

Close with:

This result follows from modeling the airfoil as a vortex sheet along the camber line.

B. Finite Wing / Induced Effects Version

Opening:

Using lifting-line theory, which accounts for induced downwash on a finite wing…

Equation:

$$\alpha = \alpha_0 - \alpha_i$$

Close with:

The reduction in effective angle of attack explains the lower lift curve slope relative to the two-dimensional case.

C. Stability Version

Opening:

Considering longitudinal static stability, which depends on the variation of pitching moment with angle of attack…

Equation:

$$\frac{\partial c_{m_{cg}}}{\partial \alpha_a} < 0$$

Close with:

A negative slope ensures that a disturbance produces a restoring pitching moment.

5-Line Emergency Version (when time is brutal)

1. Model + assumptions
2. One governing equation
3. One substitution
4. One numerical result
5. One physical sentence

That's it. Full credit *possible* even under pressure.

Why this works (grader-side translation)

Grader sees:

- ✓ Model identified
- ✓ Assumptions acknowledged
- ✓ Correct governing physics
- ✓ Logical application
- ✓ Physical understanding

Even if arithmetic is off, **method points survive**.