Higher Engineering Education Transformation

Vol.1, No.2, p. 24-32, 2026

heet.yandypress.com

Original article

Evidence-governed artificial intelligence learning in Applied Optics

Jiahong Jin

School of Physics and Electronic Science, Hunan Institute of Science and Technology, Yueyang 414006, P. R. China

Keywords:
Applied Optics
generative AI
human-AI collaboration
project-based learning
engineering problem
evidence governance

Cited as:
Jin, J., Evidence-governed artificial intelligence learning in Applied Optics. Higher Engineering Education Transformation, 2026, 1(2): 2432.
https://doi.org/10.46690/heet.2026.02.02

Abstract: Generative artificial intelligence can broaden design alternatives in engineering education, but it can also displace the reasoning and evidential judgment students are expected to develop. This study reports a practice-oriented design-based redesign of a 56-hour Applied Optics course. The course integrates a knowledge graph, project-based learning, professional optical-design software, virtual or physical experimentation, and generative artificial intelligence through an evidence-governed generate-verify-account mechanism. Students first form an initial problem representation, then use artificial intelligence to expand candidate solutions, verify key claims through disciplinary evidence such as optical theory, Ansys Zemax OpticStudio (Zemax) simulation, experiment, or traceable technical sources, and finally defend an accountable engineering decision. Implementation evidence was drawn from two successive cohorts, aggregated platform records, course-outcome attainment, and project artifacts. The final-examination failure rate decreased from 51.4% to 36.7%, while attainment of outcomes related to problem solving, research, and modern tools increased modestly; overall attainment remained approximately 0.81. Because cohorts, examinations, assessment structures, and the degree of artificial intelligence integration were not equivalent, these differences are interpreted descriptively rather than causally. The contribution is an operational framework for using artificial intelligence to expand the candidate space without lowering the evidence threshold for engineering conclusions or transferring final responsibility away from students.

1. Introduction

Generative artificial intelligence (GenAI) has shifted digital transformation in higher education from a question of access to a question of judgment. Large language models can rapidly generate explanations, calculations, code, design suggestions, and reports, but fluency is not evidence of technical validity. In engineering education, students must learn not only to produce solutions but also to justify assumptions, evaluate competing constraints, select appropriate evidence, and remain responsible for decisions. Systematic reviews of artificial intelligence (AI) in education similarly identify substantial opportunities for learning and teaching alongside persistent challenges involving assessment, ethics, and educator roles (Zawacki-Richter et al., 2019; Chiu et al., 2023). Recent work therefore emphasizes metacognitive control, corroboration, epistemic agency, and human responsibility rather than simple adoption of AI tools (Kasneci et al., 2023; Atchley et al., 2024; Tate et al., 2025; Yan et al., 2025).

This tension is sharper in complex engineering problems, which involve non-obvious solutions, incomplete information, analysis and investigation, and multiple potentially conflicting requirements (IEA, 2021). Project-based learning (PBL) can place students in sustained cycles of inquiry, design, feedback, and public explanation, but its value depends on whether activities and assessment reveal the reasoning behind the artifact rather than only the final product (Chen and Yang, 2019; Kokotsaki et al., 2016; Guo et al., 2020). GenAI complicates this requirement because it can widen the solution space while also making plausible but weakly justified answers easier to accept. Early analyses of ChatGPT in education similarly highlight both its generative potential and risks involving academic integrity, contextual reliability, and higher-order thinking (Farrokhnia et al., 2024; Tlili et al., 2023).

Research on human-AI learning collaboration is increasingly moving from efficiency toward regulation and agency. Students' GenAI use varies with task fit and expected benefit (Dai, 2025), productive collaboration requires active regulation rather than passive acceptance (Shen et al., 2025; Jou and Kao, 2026), and recent design principles recommend prior human thinking, cognitive friction, and independent corroboration (Tate et al., 2025; Vendrell and Johnston, 2026). What remains less explicit for engineering education is how these principles can be translated into an operational evidence system: AI output must have a defined epistemic status, verification must rely on

Yandy Scientific Press

∗Corresponding author.
E-mail address: jissky1@163.com
Received: June 26, 2026; revised: July 22, 2026; accepted: August 2, 2026; published: August 4, 2026.
ISSN: 3136-5221

Table 1 Implementation timeline and changing role of GenAI

Period/cohort	Main course elements	Role of GenAI	Interpretive status
Foundation before 2023-2024	Blended platform, digital resources, question bank, interaction data, virtual/simulation cases	AI-enabled resources existed, but no fixed course-wide GenAI protocol	Digital foundation; not a comparison cohort
2023-2024 (n = 37)	Project-driven teaching strengthened; telescope/lens cases; Zemax verification; additional problem sessions	GenAI not implemented as a uniform, separately identifiable intervention	Earlier reform cohort; final exam 50%
2024-2025 (n = 30)	Knowledge-graph/digital-textbook resources expanded; inquiry learning, experimental design, virtual experiment and design projects strengthened	GenAI used more explicitly for explanation and candidate expansion; technical suggestions required independent checking	Later reform cohort; final exam 40% and process evidence expanded

discipline-recognized evidence, and responsibility for the final technical judgment must remain identifiable.

Applied Optics provides a suitable setting because geometric-optics theory, professional optical-design software, virtual or physical experiments, and traceable technical sources can serve as partially independent evidence channels (de Jong et al., 2013). Building on an existing blended-learning and simulation foundation, this study therefore examines how a knowledge graph, GenAI, Zemax, and experiment can be orchestrated as a coherent problem-knowledge-candidate-evidence-decision flow; how a generate-verify-account mechanism can let AI expand design alternatives without displacing students' evidential judgment and engineering responsibility; and what implementtation evidence from two successive cohorts supports about feasibility and competence development while keeping causal claims within the limits of the design.

The study contributes an evidence-governed human-AI PBL model in which AI output is explicitly provisional, consequential technical claims must cross a discipline-recognized evidence gate, and acceptance, rejection, or revision of AI suggestions remains traceable to student reasoning. The intended contribution is therefore not a claim that AI itself improves achievement, but a course-level architecture for using generative capacity without lowering the evidence threshold for engineering conclusions or transferring final accountability away from the learner.

2. Course context and research design

2.1 Course context and iterative redesign

Applied Optics is a three-credit core professional foundation course for Optoelectronic Information Science and Engineering and related programs. It comprises 48 lecture hours and 8 practical hours and covers geometrical optics, ideal imaging systems, stops and pupils, typical optical instruments, aberrations, image-quality evaluation, photometry, and introductory optical design. Since 2022, the course has used learning platforms for pre-class resources, classroom interaction, assignments, and learning records; digital teaching materials, question banks, virtual experiments, and Zemax-based cases had also accumulated before the current redesign.

The study adopts a practice-oriented design-based research perspective (The Design-Based Research Collective, 2003). Course difficulties were diagnosed from routine teaching, translated into design requirements using relevant learning theories, implemented in successive offerings, and revised using assessment and project evidence. The reform therefore accumulated rather than appearing as a single treatment. Table 1 makes the chronology explicit and prevents the earlier cohort from being misread as an untreated control group.

2.2 Evidence sources and analytic boundaries

Evidence sources included course syllabi and project briefs, aggregated learning-platform records, course-outcome attainment forms, assessment-quality reports, project artifacts and simulation reports, and instructor observation and reflection. Routine project submissions were inspected to determine whether redesigned tasks made requirement decomposition, candidate comparison, verification, and trade-off reasoning visible. Because these artifacts were not analyzed with a common student-level qualitative coding framework, they are used as illustrative implementation evidence rather than as a formally coded outcome dataset.

The cross-cohort comparison is descriptive. The cohorts were not randomly assigned, examinations were not standardized equivalent tests, and assessment changed between years. GenAI and knowledge-graph resources were also introduced progressively rather than as a fixed intervention. Accordingly, score and outcome-attainment differences are used to describe implementation trends, not to estimate a treatment effect. Only aggregate or de-identified records are reported. Informed consent was obtained from all participating students for use of de-identified course records, project artifacts, and AI-use records for research and publication.

3. Design rationale and the evidence-governed model

3.1 Design requirements

Four recurring difficulties motivated the redesign. First, chapter-based organization made relationships among stops, aberrations, image quality, and typical systems difficult to transfer to unfamiliar tasks. A knowledge graph was therefore used to externalize prerequisite, hierarchical, attribute, and application relations; its purpose was conditional

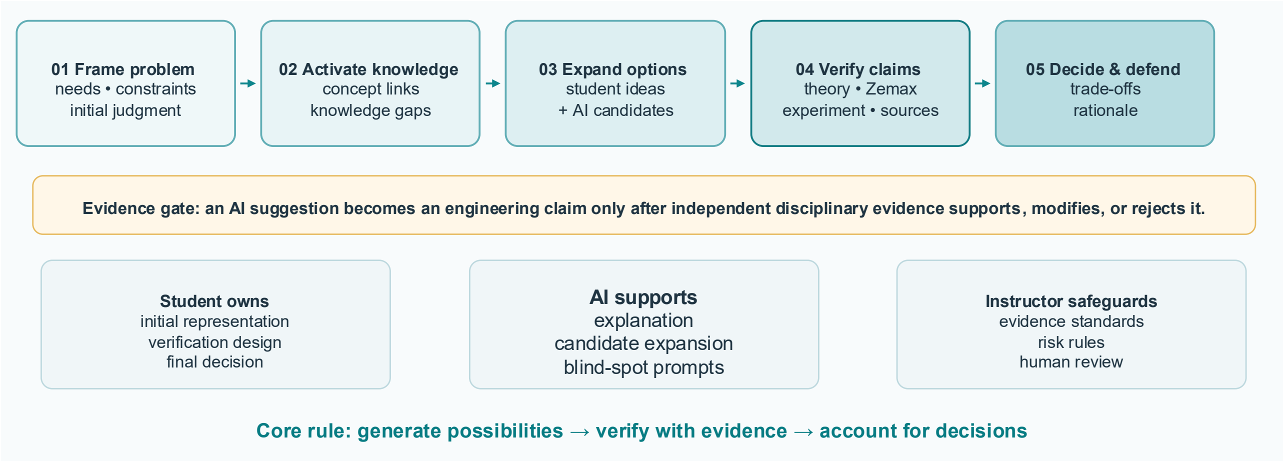

Figure 1 Course-level evidence-governed human-AI project-based learning framework

knowledge rather than visual decoration (Nesbit and Adesope, 2006). Second, conventional verification experiments underrepresented open-ended engineering trade-offs. PBL was used to require students to interpret requirements, compare alternatives, iterate, and defend a product under constraints. Third, GenAI reduced the cost of generating alternatives but created a risk of cognitive off-loading. Human-first reasoning and independent disciplinary verification were therefore built into the workflow. Fourth, platform metrics such as attendance and clicks were insufficient evidence of learning; assessment was shifted toward calculations, model versions, design rationales, verification records, and oral defense.

These responses are aligned through constructive alignment: learning outcomes are expressed as observable engineering performances, activities are selected to elicit those performances, and assessment collects evidence that can guide subsequent teaching (Biggs, 1996; Black and Wiliam, 1998). The redesigned outcomes combine four dimensions within the same project: disciplinary understanding, engineering problem solving, AI literacy and metacognition, and professional responsibility. Students must not only achieve a technically plausible design but also explain why a method applies, what evidence changed their judgment, where AI contributed, and what trade-offs remain.

PBL supplies the temporal structure for this alignment. Its role is not to add a comprehensive project after all theory has been taught, but to alternate disciplinary learning with engineering decisions. When a project exposes a new performance conflict, students return to relevant knowledge nodes, refine their understanding, and use that knowledge in the next design iteration. The knowledge graph therefore functions as a task-indexed bridge between engineering needs and disciplinary concepts: students can start from a project problem to locate the knowledge needed for the next decision, or from a knowledge node to identify where and why it is used. This preserves the chapter structure of the course while making conditional knowledge and cross-topic transfer more explicit.

Human-AI collaboration is organized around epistemic responsibility rather than efficiency-based task division. AI may organize language, expand alternatives, or surface blind spots, but it does not confer credibility on its own output. Students retain responsibility for problem representation, initial judgment, verification design, and final engineering conclusions, while instructors define evidence standards, set AI-use boundaries, and review high-risk decisions. Formative assessment closes the loop: diagnostic questions, design reviews, model versions, verification records, AI-use traces, and oral defense make it possible to identify whether difficulty lies in requirement interpretation, knowledge selection, parameter setting, simulation interpretation, or trade-off reasoning, so that evidence from one iteration can guide the next instructional response (Black and Wiliam, 1998).

3.2 Evidence governance: generate, verify, and account

Evidence governance refers to rules that determine how an AI-generated proposition may enter an engineering conclusion. In the generation phase, students first record their own problem representation, initial calculation, or structural hypothesis and then use AI to expand materials, variables, optimization directions, comparison dimensions, or possible blind spots. AI output is treated as a provisional candidate rather than submission-ready knowledge. In the verification phase, a key claim cannot be validated by asking the same model whether its earlier answer is correct; it must be checked through a disciplinary channel with at least partial independence from the generated wording, such as optical theory, paraxial calculation, Zemax simulation, virtual or physical experiment, an engineering standard, or a traceable technical source. In the accountability phase, students explain why a candidate is accepted, rejected, or modified and defend trade-offs among performance, structure, cost, tolerance, and manufacturability.

This creates a deliberate burden of verification. AI can reduce the time required to generate alternatives, but it does not lower the evidentiary threshold for accepting an engineering claim. More candidates may therefore create more evaluative work: students must decide which ideas are worth testing, select appropriate evidence, interpret disagreement, and judge when evidence is sufficient. The model uses AI's efficiency for exploration while preserving the cognitive work of engineering judgment and human responsibility emphasized in recent AI-education frameworks (Miao and Holmes, 2023; Miao et al., 2024; Yan et al., 2025).

Table 2 Five-stage implementation, human-AI division of labor, and learning evidence

Stage	Student responsibility	Permitted AI role	Instructor role	Core evidence
1. Problem framing	Interpret needs, constraints, conflicts, and missing information	Terminology explanation; checklist suggestions	Maintain authenticity; probe constraint conflicts	Requirement-constraint table; unresolved questions
2. Knowledge activation	Locate knowledge gaps and explain why selected concepts matter	Node/resource recommendations; self-check questions	Diagnose misconceptions; provide targeted theory	Knowledge-graph path; key calculations; correction record
3. Candidate generation	Produce an initial idea, then compare alternatives	Expand materials, variables, strategies, and blind spots	Ensure human-first reasoning; challenge candidate value	Initial design; candidate ledger; planned verification
4. Independent verification	Test claims with theory, Zemax, experiment, standards, or reliable sources	Explain tools or help organize results; no self-confirmation	Set evidence standards; review model settings and uncertainty	Calculation/simulation/experiment; evidence conflict record
5. Accountable decision & iteration	Accept, reject, or modify candidates; defend trade-offs and revisions	Language support and low-risk reflection	Question sufficiency of evidence; conduct oral defense	Version history; rationale; AI-use statement; final defense

The resulting asymmetry is intentional. Generation expands the possibility space, verification changes the credibility of a claim, and accountability determines whether that claim may enter the final design. Evidence conflict is not treated as failure: disagreement among theory, simulation, and experiment sends students back to assumptions, requirements, parameter choices, model settings, measurement uncertainty, or idealization. In this way, the model preserves iteration while preventing repeated assurance from the same AI model from being mistaken for independent verification. Figure 1 presents the course-level architecture of the evidence-governed human-AI PBL framework. It defines the recurring reasoning stages, the evidence gate, and the respective responsibilities of students, AI, and instructors across different course projects.

3.3 Five-stage project cycle

The model is implemented through five recurring stages: problem framing, knowledge activation, candidate generation, independent verification, and accountable decision and iteration. The knowledge graph supports retrieval and transfer, AI supports low-risk explanation and candidate expansion, professional simulation and experiments function as evidence gates, and portfolios and oral defense make reasoning traceable. Table 2 summarizes the division of labor and the evidence expected at each stage.

Across the five stages, the required evidence becomes progressively stronger. Early outputs make needs, constraints, knowledge gaps, and initial hypotheses visible; candidate records then preserve the source, expected mechanism, risk, and planned verification of important AI suggestions. Verification must disclose enough information to be auditable: calculations state assumptions and applicability, simulations report key model settings and evaluation criteria, and experimental or virtual evidence addresses uncertainty and departures from idealized conditions. The final portfolio preserves version differences, reasons for modification, and an AI-use statement so that the engineering conclusion can be traced back through the evidence chain.

Assessment follows the same logic of evidential traceability. The final examination remains important for foundational knowledge and integrated problem solving, but process evidence is used to capture performances that a time-bounded test cannot show directly. During the reform, the examination weight was reduced from 50% to 40% as inquiry learning, experimental design, and virtual experiment became more explicit in the assessment structure. Platform behavior is treated only as a diagnostic clue; stronger evidence comes from requirement analysis, key calculations, model versions, simulation or experimental results, peer and instructor feedback, acceptance/rejection rationales, and oral defense. This also gives educational value to an unsuccessful design when students can identify why it failed and justify a defensible revision.

AI use is also classified by risk. Concept explanation, resource suggestions, software-operation prompts, and language editing are low-risk uses that require normal source awareness. Parameter suggestions, design comparisons, data-processing ideas, and optimization strategies are medium risk; they must remain traceable as candidates and be checked through an independent disciplinary evidence channel before influencing the design. High-risk uses are prohibited, including generating or altering raw experimental data, fabricating simulation results, replacing the student's key derivation or oral defense, and uploading sensitive or unauthorized engineering information. Repeated agreement from the same AI model is never treated as evidence for a consequential technical claim, and academic-integrity judgments and final grading remain subject to instructor review.

4. Project implementation: preliminary smartphone-lens design

Table 3 Mapping the smartphone-lens project to selected IEA characteristics of complex engineering problems

IEA characteristic	Manifestation in the project	Observable student evidence
WP1 - Depth of knowledge required	Lens design requires application of geometric optics, aberration theory, optical materials, tolerance analysis, and engineering modelling.	Optical calculations; theoretical justification; model assumptions
WP2 - Range of conflicting requirements	Image quality, field of view, system length, material, cost, tolerance, and manufacturability impose competing requirements.	Requirement-constraint matrix; trade-off analysis
WP3 - Depth of analysis required	No single obvious lens configuration satisfies all requirements; alternative structures must be analysed and optimized.	Alternative designs; optimization records; selection/rejection rationale
WP4 - Familiarity of issues	The specific combination of imaging requirements and constraints requires students to address design situations not directly reproduced from textbook examples.	Assumption record; design iterations; justification of modelling choices
WP7 - Interdependence	Lens parameters, aberrations, packaging, materials, tolerances, and image performance are strongly coupled and require a system-level approach.	Sensitivity analysis; parameter interactions; final system-level design

4.1 Complex engineering design task

The smartphone-lens project is a course-scale analogue of a complex engineering problem rather than a full industrial lens-development task. Core requirements include focal length, field of view, total track length, and image quality, while material choice, structure, cost, tolerance, and manufacturability create additional decision dimensions. Students begin with paraxial theory to estimate focal length, aperture, and image position and to justify why an initial structure is physically plausible. This human-first step prevents blind parameter tuning and provides a baseline against which later AI suggestions can be evaluated. Table 3 shows how the task instantiates selected IEA characteristics.

4.2 Generate-verify-account in the project

After the initial design, students may use GenAI to suggest optimization variables, image-quality metrics, candidate materials, structural adjustments, or possible blind spots. Natural-language suggestions must be translated into testable engineering hypotheses. A proposal to change a surface curvature, air gap, or glass, for example, must be accompanied by an expected optical effect, possible trade-offs, and a planned verification method. Each team records the source, expected mechanism, initial judgment, verification channel, and final disposition of important candidates.

Students then construct and revise models in Zemax and examine how curvature, spacing, material, and stop position affect spherical aberration, coma, field curvature, and integrated image-quality measures. Paraxial theory, simulation, and virtual or physical experiments operate as partially independent evidence gates. When evidence sources disagree, students examine assumptions, model settings, measurement uncertainty, manufacturing or alignment error, and idealization rather than selecting whichever result is most convenient. An AI candidate enters the final design only after this evidence supports, modifies, or rejects it.

Final decisions are not ranked by a single image-quality value. Teams must defend trade-offs among optical performance, total length, material, cost, tolerance, and alignment feasibility and explain which candidates were rejected and why. Oral-defense questions focus on the basis of decisions, such as what evidence justified a variable choice or how a tighter length constraint would change the solution. The aim is to replace 'the AI suggested this' with a claim the student can defend. The same logic is used at different levels in telescope design, telecentric imaging, and other course projects, although the smartphone-lens task provides the clearest integrated example.

The smartphone-lens task is the most integrated example, but it is embedded in a layered project portfolio rather than used as an isolated capstone. Foundational work on thin-lens focal-length measurement and optical alignment emphasizes paraxial imaging, error analysis, and operational discipline; telescope design connects magnification, stops, resolution, modeling, assembly, and oral defense; telecentric imaging adds depth of field, aberrations, sensitivity, and measurement stability; and infrared or laser-system scenarios require students to combine simulation, technical sources, and risk analysis. Across these levels, the role of AI changes with task risk: it may support preparation and explanation in foundational activities, broaden comparison strategies in integrative tasks, and organize information in innovative scenarios, while consequential technical claims remain subject to independent evidence. This progression shows that generate-verify-account is intended as a course-level reasoning rule, not a workflow tied only to one lens-design problem. Figure 2 operationalizes the course-level framework in Figure 1 for the smartphone-lens project by translating the generic stages into lens-specific design tasks, verification channels, and accountable outputs.

5. Implementation evidence and discussion

5.1 Cohort evidence

Aggregated platform records across 2023-2025 show an average classroom attendance rate of 97%, a median of 100%, and an in-class response rate of approximately 90%,

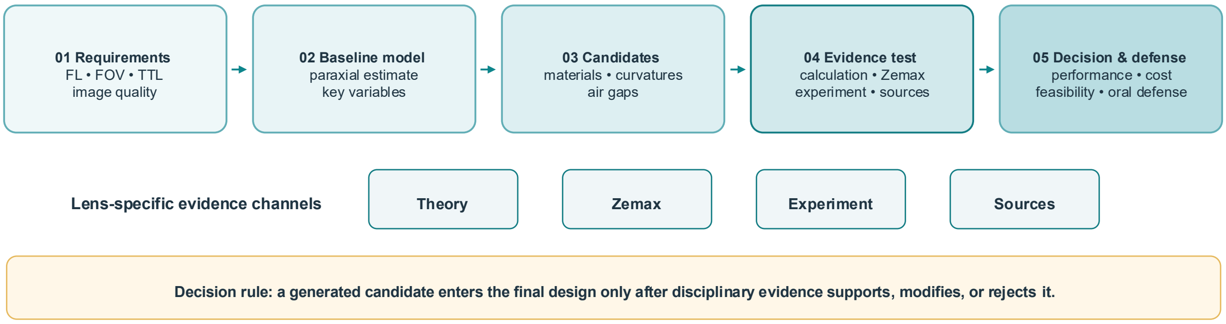

Figure 2 Project-level implementation of the evidence-governed framework in the smartphone-lens project

Table 4 Descriptive comparison of two successive course cohorts

Indicator	2023-2024 (n = 37)	2024-2025 (n = 30)	Descriptive change
Final-examination mean (SD)	59.6 (10.73)	61.9 (11.97)	+2.3 points
Final-examination failure rate	51.4% (19/37)	36.7% (11/30)	−14.7 percentage points
Overall course-grade mean (SD)	77.1 (5.56)	78.9 (5.77)	+1.8 points
Course grade ≥80	24.3% (9/37)	56.7% (17/30)	+32.4 percentage points
Course Outcome 1 (Optical Fundamentals)	0.757	0.750	−0.007
Course Outcome 2 (Problem-Solving Skills)	0.795	0.820	+0.025
Course Outcome 3 (Practical Application)	0.710	0.740	+0.030
Course Outcome 4 (Values and Self-Learning)	0.969	0.920	−0.049; assessment basis changed
Overall attainment	0.808	0.810	Essentially stable

indicating a stable participation base but not cognitive engagement. More direct outcome evidence is available for the two successive cohorts. Table 4 summarizes the descriptive comparison. The final-examination mean changed from 59.6 to 61.9 and the examination failure rate from 51.4% to 36.7%; the overall course-grade mean changed from 77.1 to 78.9 and the proportion scoring 80 or above increased. Course Outcomes 2 and 3, which include problem solving, research, lifelong learning, and use of modern tools, increased modestly, while Outcome 1 was stable and Outcome 4 remained high but was assessed differently. Overall attainment was essentially unchanged.

Note. Course-grade distributions are reported descriptively because assessment structures and component weights differed between cohorts.

The overall course-grade mean changed only modestly, from 77.1 to 78.9, whereas the proportion of students scoring 80 or above increased from 24.3% to 56.7%. These two indicators need not change proportionally because the latter is sensitive to the distribution of scores around a fixed cutoff. More importantly, the course-grade composition changed between the two cohorts: the final-examination weight decreased from 50% to 40%, while inquiry learning, experimental design, virtual experimentation, and other process-based evidence became more explicit in the assessment structure. The available aggregate data do not allow the observed shift around the 80-point threshold to be separated from these assessment changes. Therefore, the increase in the proportion scoring 80 or above is interpreted only as a descriptive change in the observed grade distribution, rather than as direct evidence of improved learning outcomes.

Project artifacts provide evidence closer to the proposed mechanism than attendance or aggregate grades, although they were not coded as a formal student-level outcome dataset. Routine submissions included theoretical estimates, model files, parameter versions, image-quality evaluations, virtual or physical verification, records of accepted or rejected AI suggestions when available, and reasons for engineering trade-offs. Instructor review correspondingly shifted from checking whether software ran or whether an output resembled a reference value toward asking why a variable was selected, whether evidence sources were mutually consistent, and how the design should respond if constraints changed. These records show that the redesign can make previously implicit reasoning public and assessable; they do not quantify growth in that reasoning or demonstrate that the generate-verify-account pathway caused the cohort-level differences.

At the same time, the 36.7% final-examination failure rate in 2024–2025 remains an important negative result, indicating that foundational understanding and transfer under

Table 5 Evidence sources, supported inferences, and interpretive limits

Evidence source	Main observation	Inference supported	Limit
Platform records	High attendance and response rates	Stable participation base for blended/interactive learning	Behavioral participation is not cognitive engagement and cannot be attributed to AI
Projects and simulation reports	Submissions contain calculations, model versions, verification evidence, and trade-off explanations	The redesign can make reasoning and verification visible and assessable	Artifacts were not coded with a common student-level scheme; growth is not quantified
Course-outcome records	Outcomes 2 and 3 increased; Outcome 1 stable; overall ≈0.81	Directional shift is consistent with stronger emphasis on inquiry, design, and modern tools	Cohorts, exams, task weights, and evidence bases were non-equivalent
Instructor review/oral defense	Questions shifted toward reasons, evidence sufficiency, and response to changed constraints	Generate-verify-account can be enacted within normal project teaching	Teacher observation is not an independent causal measure

individual examination conditions remain unresolved despite higher values in some broader course-outcome indicators.

The outcome profile was also non-uniform. Outcome 1 remained essentially stable (0.757 to 0.750), whereas Outcomes 2 and 3 increased modestly from 0.795 to 0.820 and from 0.710 to 0.740, respectively. Outcome 4 decreased from 0.969 to 0.920 but remained high, with a changed assessment basis, while overall attainment remained approximately 0.81. These patterns indicate a directional shift in the observed competence profile rather than a general improvement across outcomes.

5.2 Interpreting the evidence

Table 5 summarizes the main evidence sources and the limits of the inferences that can be drawn from them. Taken together, the evidence is more informative about how the redesigned course operated in practice than about whether the redesign produced a measurable treatment effect. The evidence mainly shows that the redesigned course can make students' reasoning, verification, and decision-making more visible. The knowledge graph supports knowledge retrieval, AI broadens possible solutions, and theory, simulation, experiment, and technical sources provide evidence for evaluating those solutions. Project records and oral defense further require students to explain why a design was accepted, revised, or rejected.

The cohort results should be interpreted cautiously. The lower examination failure rate and modest increases in Outcomes 2 and 3 are consistent with the stronger emphasis on inquiry, design, and modern tools. However, Outcome 1 remained stable, overall attainment changed little, and the cohorts and assessment structures were not equivalent. The results therefore support feasibility and directional change rather than a causal effect of GenAI or the redesign.

5.3 Contribution and transferability

The main contribution of this study is to connect AI use with disciplinary evidence and student responsibility. AI may broaden possible solutions, but important technical claims must still be checked through calculations, simulation, experiment, standards, or other reliable sources, and final decisions remain with students. This principle can extend beyond optical design to other engineering courses in which technical claims can be independently verified. What is transferable is therefore not a specific tool, but the sequence of generating possibilities, verifying them with disciplinary evidence, and making accountable engineering decisions.

5.4 Limitations and future work

The study nevertheless has clear limitations. The two cohorts are modest in size (37 and 30 students), examinations and assessment weights were not equivalent, and knowledge-graph resources, PBL, virtual simulation, and GenAI accumulated across the reform, so their separate contributions cannot be identified. Project versions, selected AI-use traces, and instructor observations were not analyzed through a common student-level coding framework, and the course did not archive a complete model-by-model profile of GenAI use. The work was also conducted in one course at one institution. These constraints mean that the proposed candidate-generation, evidence-selection, conflict-resolution, and accountable-decision pathway is operationally visible but not directly validated as a causal learning mechanism.

Future validation should therefore use parallel classes or a common pre/post design task with shared rubrics and student-level process data, report effect sizes and confidence intervals, and systematically code student-AI dialogue, candidate ledgers, model versions, verification quality, and accountable decisions. Such analysis could test whether candidate diversity, evidence selection, conflict resolution, and decision quality form a pathway between AI use and complex engineering problem solving, and whether that pathway differs with prior preparation. The next empirical question is thus not simply whether students score higher when AI is available, but which human-AI collaboration pathways, under which evidence rules, support credible engineering judgment for different learners.

6. Conclusion

The central challenge of GenAI integration in engineering education is not access to another intelligent tool but the risk that inexpensive generation will be mistaken for credible engineering knowledge. The Applied Optics redesign addresses this risk through an evidence-governed project-based sequence of generate, verify, and account. Students form an initial problem representation, use AI to expand candidate solutions, test key claims against optical theory, professional simulation, experiment, or reliable technical sources, and remain responsible for the final decision and trade-off.

Implementation evidence from two successive cohorts supports the feasibility of the model and shows descriptive changes consistent with stronger emphasis on problem solving, research, and modern tools, while also revealing persistent weaknesses in individual examination performance. Because the reform evolved across cohorts and assessment changed, these observations are not treatment-effect estimates. The contribution is therefore the instructional architecture and its explicit epistemic boundary: AI may expand possibilities, but engineering conclusions require independent evidence and accountable human judgment.

Acknowledgments

This work was supported by the Hunan Provincial Teaching Reform Research Project for Higher Education (No. HNJG-20230854), and the Smart Course Construction Project of Hunan Institute of Science and Technology, “Smart Course Construction for Applied Optics.”

Conflicts of interest

The authors declare no competing interests.

Open Access

References

Atchley, P., Pannell, H., Wofford, K., et al. Human and AI collaboration in the higher education environment: Opportunities and concerns. Cognitive Research: Principles and Implications, 2024, 9(1): 20.

Biggs, J. Enhancing teaching through constructive alignment. Higher Education, 1996, 32(3): 347-364.

Black, P., Wiliam, D. Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 1998, 5(1): 7-74.

Chen, C. H., Yang, Y. C. Revisiting the effects of project-based learning on students’ academic achievement: A meta-analysis investigating moderators. Educational Research Review, 2019, 26: 71-81.

Chiu, T. K. F., Xia, Q., Zhou, X., et al. Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. Computers and Education: Artificial Intelligence, 2023, 4: 100118.

Dai, Y. Why students use or not use generative AI: Student conceptions, concerns, and implications for engineering education. Digital Engineering, 2025, 4: 100019.

de Jong, T., Linn, M. C., Zacharia, Z. C. Physical and virtual laboratories in science and engineering education. Science, 2013, 340(6130): 305-308.

The Design-Based Research Collective. Design-based research: An emerging paradigm for educational inquiry. Educational Researcher, 2003, 32(1): 5-8.

Farrokhnia, M., Banihashem, S. K., Noroozi, O., et al. A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations in Education and Teaching International, 2024, 61(3): 460-474.

Guo, P., Saab, N., Post, L. S., et al. A review of project-based learning in higher education: Student outcomes and measures. International Journal of Educational Research, 2020, 102: 101586.

International Engineering Alliance (IEA). Graduate attributes and professional competences: Approved version 4, 21 June 2021.

Jou, M., Kao, C. C. Reframing human-AI collaboration in higher education: Behavioral pathways from generative AI to critical thinking and ethical creativity. Journal of Computing in Higher Education, 2026: 1-30.

Kasneci, E., Seßler, K., Küchemann, S., et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 2023, 103: 102274.

Kokotsaki, D., Menzies, V., Wiggins, A. Project-based learning: A review of the literature. Improving Schools, 2016, 19(3): 267-277.

Miao, F., Holmes, W. Guidance for generative AI in education and research. Paris: UNESCO, 2023.

Miao, F., Shiohira, K., Lao, N. AI competency framework for students. Paris: UNESCO, 2024.

Nesbit, J. C., Adesope, O. O. Learning with concept and knowledge maps: A meta-analysis. Review of Educational Research, 2006, 76(3): 413-448.

Shen, W., Lin, X. F., Liu, J., et al. The process of undergraduates' collaboration with a generative artificial intelligence chatbot: Insights from conversation content and epistemic network analysis. Journal of Computer Assisted Learning, 2025, 41(6): e70140.

Tate, T. P., Harnick-Shapiro, B., Ritchie, D. R., et al. Incorporating generative AI into a writing-intensive undergraduate course without off-loading learning. Discover Computing, 2025, 28(1): 72.

Tlili, A., Shehata, B., Adarkwah, M. A., et al. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 2023, 10: 15.

Vendrell, M., Johnston, S. K. Scaffolding critical thinking with generative AI: Design principles for integrating large language models in higher education. Computers and Education: Artificial Intelligence, 2026, 10: 100572.

Yan, L., Pammer-Schindler, V., Mills, C., et al. Beyond efficiency: Empirical insights on generative AI's impact on cognition, metacognition and epistemic agency in learning. British Journal of Educational Technology, 2025, 56(5): 1675-1685.

Zawacki-Richter, O., Marín, V. I., Bond, M., et al. Systematic review of research on artificial intelligence applications in higher education - where are the educators? International Journal of Educational Technology in Higher Education, 2019, 16: 39.