Article

Development of a Multidimensional Evaluation Indicator System and Evidence Mapping for the Training Quality of Professional Master's Students in Mechanical Engineering: An Outcome-Based Approach

Yi Zhang[1,*], Ankang Zhu[1], Jisheng Ma[1] and Yongxing Li[1]

[1] School of Mechanical Engineering, Xijing University, Xi'an, China
* Correspondence: Yi Zhang, School of Mechanical Engineering, Xijing University, Xi'an, China

Abstract: To address challenges in evaluating the training quality of professional master's students in Mechanical Engineering, including unclear relationships between research outputs and demonstrated competencies, the potential for double-counting the same technical work, and the risk of drawing unwarranted conclusions from insufficient evidence, this study develops an analytical framework grounded in outcome-based education (OBE), integrating the principles of constructive alignment, engineering education competency frameworks, and assessment validity. The proposed framework follows a five-stage analytical pathway: training objectives–core competencies–evaluation indicators–performance evidence–quality judgments. Using normative document analysis, directed content analysis, and retrospective multiple-case comparison, the study establishes a competency evaluation system comprising six first-level dimensions and twelve second-level indicators. It further specifies behaviorally anchored assessment rubrics, rules for mapping multiple sources of evidence to competency indicators, and procedures for avoiding duplicate recognition of technical tasks. Evidence mapping was conducted using nine mechanical engineering research and development cases drawn from a case repository developed at Xijing University in 2021. The findings indicate that different forms of research outputs may either originate from the same core technical task or reflect distinct engineering activities. Existing case materials provide relatively accessible evidence of design, implementation, and certain verification activities. However, judgments concerning technical quality, individual contributions, and fulfillment of ethical and safety responsibilities require additional support from primary data and process documentation. The principal contribution of this study lies in establishing explicit rules for linking evaluation indicators to engineering evidence and defining the interpretive boundaries of competency-related judgments, rather than empirically measuring students' competency levels. The proposed framework offers a methodological reference for evidence-based evaluation of training quality in professional master's degree programs. Nevertheless, its scoring consistency, construct validity, and applicability across different educational contexts require further empirical validation.

Keywords: outcome-based education; professional master's students in Mechanical Engineering; training quality evaluation; competency indicators; multi-source evidence; case analysis

Received: 11 September 2026
Revised: 02 November 2026
Accepted: 15 November 2026
Published: 19 November 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Research Background and Problem Statement

Professional postgraduate education has a vital mission to serve industry development and cultivate highly qualified professionals with practical and innovative capabilities. As the differentiated development of academic and professional postgraduate education continues to advance, institutional arrangements, including the option of applying for degrees on the basis of practice-based achievements, have further diversified the forms in which professional degree outcomes may be presented [1-3].

However, diversification in the forms of degree-related outputs does not automatically ensure the validity of competency assessment. While institutional regulations specify acceptable types of outputs and their quality requirements, the evaluation of educational quality must still address two fundamental questions: what competencies students demonstrate in authentic engineering tasks, and whether the evidence supporting such judgments is sufficient.

Professional master's students in Mechanical Engineering may produce project reports, competition entries, technical patents, and degree theses through activities such as product development, structural design, computational simulation, prototype debugging, and performance validation. These materials may represent different records of the same technical task, or they may arise from different stages or distinct subtasks. Simply aggregating outputs according to their quantity, ranking, or type of documentation may not only result in double-counting technical contributions but also overlook individual differences in collaborative work and the influence of task complexity and resource availability on performance. Accordingly, the fundamental unit of evaluation should shift from the "output artifact" to the "specific engineering task undertaken by the student and the associated verifiable performance."

Against this background, this study addresses three interrelated questions. First, how can broadly formulated training objectives be translated into engineering competency indicators with clearly defined behavioral manifestations and quality criteria? Second, how can evidence embedded in different outputs be distinguished as originating from the same technical work, providing complementary information, or representing independent contributions, thereby avoiding duplicate recognition of identical technical activities? Third, how can the defensible scope of competency judgments be determined according to the relevance, authenticity, individual attribution, and sufficiency of the available evidence? The study evaluates students' demonstrated competencies rather than using institutional resource inputs or the quantity of research outputs as substitutes for individual educational quality.

1.2. Related Research and Positioning of the Present Study

In recent years, research on outcome-based and competency-based engineering education has gradually shifted from the description of educational objectives toward the operationalization of competency indicators and assessment through authentic tasks. Reviews of outcome-based education (OBE) in higher education have identified its major research themes [4]. Assessment research in engineering education has also been urged to pay greater attention to workplace contexts and the educational programme level [5]. Structured indicator frameworks have been proposed from the perspective of professional practical competencies [6]. The integrated development of professional and generic competencies has also been examined [7]. These studies provide a foundation for identifying competency structures. Nevertheless, the identification of competency indicators does not, in itself, establish corresponding rules for evidence-based assessment. In particular, it remains necessary to clarify which engineering activities can legitimately support judgments about specific competencies.

Regarding assessment tasks and the use of evidence, authentic engineering problems have been emphasized as a means of revealing students' technical decision-making capabilities, and different approaches to and purposes of work-integrated learning assessment have been summarized [8,9]. Authentic assessment and the development of higher-order competencies have been examined, while a lack of consistency in the conceptualization and measurement criteria of certain competencies in engineering education has also been noted [10,11]. A further review of work-integrated learning in engineering education has demonstrated the importance of establishing explicit connections between authentic contexts and assessment activities [12]. Collectively, these studies suggest that competency dimensions, task contexts, and assessment instruments must be constructively aligned. However, the evidentiary relationships among multiple forms of outputs, such as engineering projects, competition entries, and degree theses, in

supporting inferences about an individual student's competencies require further clarification.

Unlike studies that construct composite indices based on training components, output quantities, or overall satisfaction, the present study focuses on the interpretable relationships among competency indicators, technical activities, and evidence quality. On the one hand, relatively distinct competency dimensions are derived from the sequence of mechanical engineering tasks to establish observable and verifiable assessment criteria. On the other hand, specific technical tasks are adopted as the units of analysis to identify duplicate documentation across different outputs and to distinguish three levels of information: the existence of an output, the technical work actually performed, and the quality of the demonstrated competency. The study is limited to developing an indicator framework and mapping evidence from historical cases. It does not claim that the proposed assessment instrument has undergone reliability or validity testing, nor does it establish that students' competencies have improved.

2. Theoretical Foundations and Analytical Framework for Evaluation

2.1. Competency Operationalization in Outcome-Based Assessment

Outcome-based education emphasizes the organization of educational activities and assessment around intended learning outcomes, while constructive alignment requires consistency among educational objectives, learning tasks, and assessment methods [13]. For professional master's students in Mechanical Engineering, the ability to "solve complex engineering problems" represents a broadly defined educational objective that must be further operationalized into observable behaviors, such as identifying engineering requirements, evaluating and justifying technical approaches, designing solutions, implementing engineering solutions, and conducting experimental validation. The CDIO competency framework suggests that, beyond disciplinary knowledge and engineering systems competencies, interpersonal collaboration, professional standards, and social responsibility are also essential competency requirements in engineering practice [14]. Therefore, the indicator structure should encompass the principal stages of engineering problem-solving while avoiding the direct classification of different forms of outputs as distinct competency dimensions.

Assessment validity concerns the extent to which inferences drawn from observed evidence to conclusions about competencies are adequately justified [15]. Accordingly, this study distinguishes three progressively more demanding levels of assessment information. Output confirmation addresses whether a particular output has been produced; identification of competency-related performance determines which technical activities a student actually undertook; and quality judgment further establishes the level at which those activities were performed under specified constraints. For example, receiving an award in a discipline competition confirms an external evaluation outcome but does not, by itself, demonstrate that a particular team member independently completed a critical design task. Only by integrating technical documentation, process records, performance evidence, and materials documenting individual contributions can a defensible chain of inference be established between observed activities and specific competency indicators.

2.2. Analytical Framework for Multidimensional Evaluation

Building on the theoretical foundations outlined above, this study develops a five-layer analytical framework comprising training objectives, core competencies, evaluation indicators, performance evidence, and quality judgments (Figure 1).

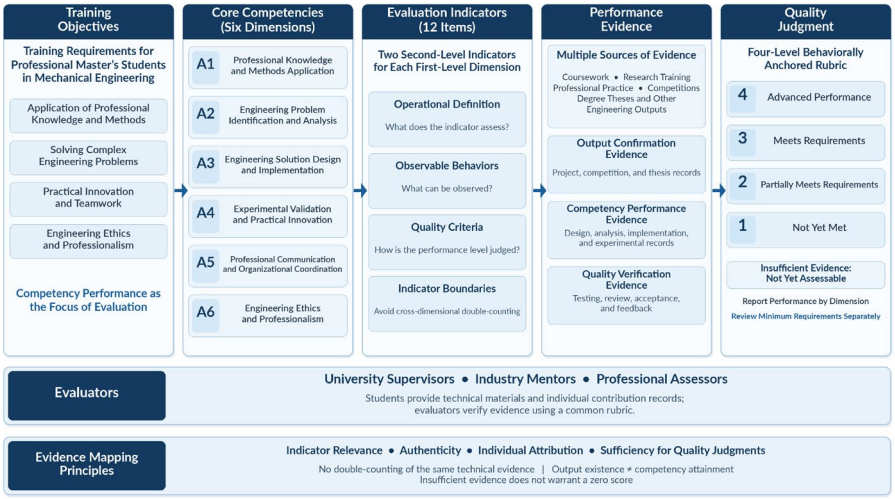


Figure 1. Analytical Framework for the Multidimensional Evaluation of Educational Quality in Professional Master's Programs in Mechanical Engineering from an Outcome-Based Perspective

Within this framework, training objectives establish the intended competency directions; core competencies provide a structured classification; evaluation indicators operationalize competencies into observable behaviors and quality criteria; and courses, research projects, professional practice, competitions, and degree-related outputs serve as different forms of evidence. Academic supervisors, industry mentors, and professional assessors then make evidence-supported judgments according to shared criteria. This approach emphasizes bidirectional verification between competency objectives and sources of evidence: objectives determine what evidence needs to be collected, while the sufficiency of the available evidence constrains the conclusions that can reasonably be drawn.

Within this framework, compliance with training regulations and the assessment of student competencies should be conducted separately. Course credits, completion of professional practice, and degree application requirements should be verified against the applicable institutional regulations, whereas engineering competencies should be assessed on the basis of actual task performance and behaviorally anchored rubrics. Minimum requirements relating to academic integrity, engineering safety, and professional ethics should be subject to independent compliance checks and must not be compensated for by high scores in other competency dimensions. Furthermore, when evidence is missing, a distinction must be made between "absence of documentation" and "failure to meet the required standard." The former indicates insufficient assessment information and should, in principle, be classified as "not yet assessable" rather than automatically assigned a low score.

3. Research Design and Methods

3.1. Data Sources and Units of Analysis

This study examines the cultivation of professional master's degree students in Mechanical Engineering at Xijing University. The 2023 edition of the Master's Degree Training Program in Mechanical Engineering, as cited in the manuscript, serves as the normative basis for developing the evaluation indicators [16]. Nine engineering research and development (R&D) cases documented in the 2021 Case Repository Development for the "Three-in-One" Training Model Integrating Research Projects, Discipline Competitions, and Degree Theses are used as historical materials for analysis. Existing statistical reports on research outputs are also consulted to identify relationships among different forms of research outcomes. Because the normative document and the case materials originate from different years, the requirements specified in the 2023 training

program are used solely to establish the evaluation framework, rather than to retrospectively assess the compliance of graduates from earlier cohorts.

The study employs two distinct units of analysis. For indicator development, the units of analysis are the training objectives and their corresponding competency requirements; for case-based evidence mapping, the units are engineering tasks or subtasks with distinct technical significance. Aggregate statistics on 67 degree theses completed between 2015 and 2020 are used exclusively to describe the relationships among research funding, discipline competitions, and degree theses, whereas the nine cases are examined to compare the evidence structures associated with different engineering tasks. These two datasets do not constitute a comprehensive student-level database and therefore cannot be combined to calculate competency scores, examine between-group differences, or estimate the causal effects of the training model. The cases were selected purposively, and the findings should not be generalized as representing the proportional distribution of characteristics across the entire postgraduate student population.

3.2. Indicator Development and Coding of Case Evidence

Drawing on a theory-driven, directed content analysis approach [17] and the principles of document analysis concerning the context of document production, source characteristics, and interpretive limitations [18], candidate indicators were developed through five sequential steps: identification of objectives, decomposition of competencies, operationalization of behaviors, mapping of evidence, and review of interpretive boundaries. Specifically, competency elements were first extracted from the training requirements and representative engineering tasks. Observable behavioral manifestations were then identified through activities such as requirements analysis, evaluation and justification of technical solutions, design and implementation, technical verification, and collaboration and fulfillment of responsibilities. Subsequently, potential sources of evidence and corresponding quality criteria were specified. Indicators were retained primarily on the basis of their relevance to training objectives, necessity for engineering tasks, observability, conceptual distinctiveness, and applicability across different forms of research outputs. The resulting framework, comprising six dimensions and twelve indicators, is a research-derived structure grounded in theoretical reasoning and document analysis. As it has not yet undergone expert assessment of content validity, it should not be characterized as a statistically validated or optimal dimensional structure.

Case analysis followed a coding sequence consisting of technical object–core problem–task activities–documented outputs–competency indicators. First, the engineering object and core technical problem were examined to determine whether different outputs originated from the same underlying technical work. Second, project, competition, and thesis records documenting substantively identical technical content were identified and consolidated, while additional experiments, design improvements, and independent subtasks were retained as separate analytical elements. Finally, each item of evidence was assessed to determine whether it could support confirmation of an output, identification of a competency-related performance, or determination of a quality level. A single technical activity could provide evidence relevant to multiple competencies when assessed against distinct evaluation criteria. For example, design decision-making and experimental verification could be mapped to indicators A31 and A41, respectively. However, the recurrence of the same activity across multiple documents did not result in repeated scoring of the same indicator. In this study, "coding" refers to the interpretation of cases and classification of evidence; it does not imply that intercoder reliability testing has been conducted.

3.3. Research Quality Control and Boundaries of Interpretation

To minimize interpretive bias, the study employed source triangulation, verification of task content, and cross-case comparison. When the same research outputs appeared in multiple reports, their technical content was examined to identify duplication. Discrepancies between the titles of records and the tasks actually undertaken were

preserved rather than artificially reconciled. Where the existence of an output could be confirmed but original design documentation, experimental records, or evidence of individual contributions was unavailable, the material was reported only as potential evidence, without assigning a performance level. In this study, "evidence mapping" refers to a conceptually reasonable correspondence between documentary materials and competency indicators; it does not establish that the corresponding competencies have attained Level 3 or Level 4. Because no formal expert-based weighting, individual student-level competency scoring, inter-rater reliability assessment, or validity testing was conducted, the conclusions are restricted to the development of the evaluation framework, the assessment of evidence coverage, and the delineation of boundaries for evaluative judgments.

4. Multidimensional Evaluation Indicator System and Assessment Rules

4.1. Indicator Structure and Operational Definitions

Guided by the principles of competency orientation, task relevance, and parsimony, a multidimensional evaluation indicator system comprising six primary dimensions and twelve secondary indicators was developed (Table 1).

Table 1. Multidimensional Evaluation Indicator System for the Educational Quality of Professional Master's Students in Mechanical Engineering

Primary dimension	Secondary indicator	Main assessment focus	Representative assessment evidence
A1 Professional Knowledge and Methodological Application	A11 Application of Disciplinary Theories	Applying disciplinary knowledge in mechanical engineering to analyze and solve technical problems	Course design assignments, technical calculations, project analysis reports
	A12 Application of Engineering Methods and Tools	Selecting and appropriately applying modeling, simulation, and technical analysis methods	Mathematical models, simulation files, technical analysis records
A2 Engineering Problem Identification and Analysis	A21 Identification of Engineering Requirements and Constraints	Identifying engineering requirements and constraints related to performance, cost, safety, and other factors	Requirements analyses, project specifications, feasibility studies
	A22 Engineering Problem Analysis and Justification	Analyzing technical problems, comparing potential solutions, and providing reasoned justification	Problem analysis reports, technical evaluations, research records
A3 Engineering Solution Design and Implementation	A31 Technical Solution Design	Developing feasible technical solutions and appropriately determining design parameters and technical approaches	Design proposals, engineering drawings, technical review comments
	A32 Engineering Implementation and System Integration	Performing structural design, assembly, commissioning, system integration, and related engineering activities	Functional prototypes, assembly records, commissioning reports
A4 Experimental	A41 Experimental Design and	Designing validation procedures, analyzing test data,	Experimental protocols, test data,

Validation and Practical Innovation	Technical Validation	and evaluating technical performance	performance analysis reports
	A42 Technical Optimization and Innovative Application	Implementing technical improvements and demonstrating innovation and practical engineering value	Optimization records, technical comparisons, application evaluations
A5 Professional Communication and Organizational Collaboration	A51 Engineering Collaboration and Individual Contribution	Undertaking clearly defined technical responsibilities and contributing effectively to teamwork	Task allocation records, work logs, supervisor evaluations
	A52 Technical Communication and Presentation of Outcomes	Clearly communicating technical solutions and analytical findings and responding to professional inquiries	Technical reports, master's theses, oral defense records
A6 Engineering Ethics and Professionalism	A61 Academic Standards and Professional Integrity	Adhering to academic ethics, intellectual property requirements, and professional codes of conduct	Original research records, compliance reviews, integrity assessments
	A62 Engineering Safety and Social Responsibility	Identifying safety, environmental, and related professional responsibilities and taking appropriate measures	Safety analyses, risk assessments, compliance records

Dimension A1, Professional Knowledge and Methodological Application, assesses the appropriate application of disciplinary knowledge and engineering tools. Dimension A2, Engineering Problem Identification and Analysis, focuses on the identification and analysis of engineering requirements, constraints, and technical problems. Dimension A3, Engineering Solution Design and Implementation, evaluates technical design decisions and system implementation. Dimension A4, Experimental Validation and Practical Innovation, addresses evidence-based technical validation and improvement. Dimension A5, Professional Communication and Organizational Collaboration, examines technical communication, teamwork, and individual contributions. Dimension A6, Engineering Ethics and Professionalism, evaluates adherence to professional standards and the fulfillment of engineering responsibilities.

The six-dimensional structure is intended to encompass both the principal stages of engineering problem-solving and the professional requirements of engineering practice, rather than treating courses, competitions, theses, or other output formats as separate evaluation dimensions. The inclusion of two secondary indicators within each dimension represents a research-informed design choice intended to balance coverage of essential competencies with the practical feasibility of the evaluation instrument. The structural adequacy of this arrangement remains subject to expert review and empirical assessment using actual performance ratings..

Although the indicators are interconnected within engineering tasks, the criteria used to assess them should remain conceptually distinct. For example, A22 focuses on technical problem analysis and the justification of alternative approaches, whereas A31 assesses the selection of technical solutions and decisions concerning design parameters. A32 concerns engineering implementation and system integration; A41 addresses experimental design, data analysis, and validation conclusions; and A42 focuses on technical improvements and their practical application value.

A single document may provide different types of evidence relevant to multiple indicators. However, the same technical contribution should not receive duplicate credit merely because it appears in several output formats. For activities spanning multiple competency dimensions, assessment should be based on the specific observable behaviors

and performance criteria associated with each indicator, rather than on the document type in which the activity is recorded.

4.2. Evidence Mapping and Deduplication Across Multiple Engineering Outputs

To avoid equating the existence of engineering outputs with competency attainment, assessment evidence is classified into three categories: output-verification evidence, competency-performance evidence, and quality-validation evidence.

Output-verification evidence primarily includes project approval records, competition awards, academic publications, and patent records. Such evidence establishes the existence and basic characteristics of an output. Competency-performance evidence comprises requirements analyses, engineering drawings, simulation records, commissioning procedures, and experimental plans, which help identify the professional activities actually undertaken by students. Quality-validation evidence includes original test data, controlled experiments, technical reviews, engineering acceptance records, and feedback from practical applications, providing a basis for judging the quality of the work performed.

These three categories should not be aggregated simply according to the number of available records. Instead, they serve distinct but related purposes: establishing what outputs were produced, identifying how the underlying engineering work was performed, and determining the quality of that performance. A defensible assessment requires consideration of four criteria: relevance to the indicator, authenticity of the evidence source, identifiability of individual contributions, and sufficiency of evidence for quality judgments. When one or more essential criteria are not met, the scope of the assessment inference must be restricted accordingly.

Deduplication across engineering outputs should be based on the underlying technical task rather than the title or number of reported outputs. Accordingly, a mapping relationship linking student, engineering task, technical evidence, and evaluation indicator should be established.

When substantially identical technical solutions are presented in project reports, competition submissions, and master's theses, the overlapping materials should be used for cross-verification rather than treated as independent contributions. Conversely, additional parameter studies, system improvements, or distinct technical subtasks may be documented separately when they provide evidence of substantively different competency performance.

For team-based outputs, individual responsibilities should be verified through work logs, version histories, supervisor assessments, or technical defenses. Digital records may facilitate evidence preservation and traceability; nevertheless, their authenticity and the limits of the conclusions drawn from them require professional scrutiny [19].

The purpose of this approach is not to diminish the value of engineering outputs, but to ensure that the unit of assessment corresponds to the student's actual competency performance.

4.3. Performance-Level Rubrics and Multi-Assessor Collaboration

To establish an observable basis for assessing competency performance, a four-level behaviorally anchored rating rubric is adopted: Level 4, Advanced Performance; Level 3, Meets Requirements; Level 2, Partially Meets Requirements; and Level 1, Does Not Yet Meet Requirements. When the available evidence is insufficient to support a defensible rating, the indicator is designated as Not Assessable Due to Insufficient Evidence.

Performance levels are determined through an integrated consideration of the student's independence in completing engineering tasks, ability to address critical engineering constraints, quality of technical justification, and adequacy of outcome validation. Competition awards, patent counts, and publication records should not be directly converted into competency scores.

Table 2 presents the performance-level rubric for A31 (Technical Solution Design), illustrating how assessment criteria distinguish basic solution feasibility from independent technical reasoning and justification. In practical applications, behaviorally

anchored criteria of comparable specificity should also be developed for the remaining eleven indicators to reduce reliance on assessors' overall impressions.

Table 2. Performance-Level Assessment Rubric for Technical Solution Design (A31)

Performance level	Assessment criteria
Level 4: Advanced Performance	Independently develops and compares alternative technical solutions, systematically considers critical engineering constraints, and provides well-supported justifications for major design decisions. The proposed solution demonstrates a high degree of technical soundness and practical feasibility.
Level 3: Meets Requirements	Develops a feasible technical solution in response to engineering requirements, appropriately selects technical approaches, and provides reasonable explanations for major design parameters and key technical choices.
Level 2: Partially Meets Requirements	Completes some technical design tasks, but consideration of engineering constraints is insufficient, and certain design decisions lack adequate justification.
Level 1: Does Not Yet Meet Requirements	The proposed technical solution fails to satisfy fundamental engineering requirements, or key design decisions lack a sound rationale, indicating that the student is not yet able to independently complete the principal design tasks.

University supervisors, industry supervisors, and technical reviewers may have access to different aspects of assessment evidence, including theoretical and methodological application, engineering implementation processes, and technical quality. Assessment responsibilities should therefore be assigned according to the evidence available to each assessor and their areas of professional expertise, rather than by simply averaging subjective impressions.

Assessors should provide independent judgments using a common rubric. Where substantial disagreements arise, the relevant technical records, experimental findings, and documentation of individual contributions should be reexamined. If sufficient evidence remains unavailable, the indicator should retain the designation Not Assessable Due to Insufficient Evidence.

Previous research has proposed approaches to assessing course and program outcome attainment using multiple sources of evidence [20]. Such work provides methodological guidance for the future implementation of the proposed framework; however, it should not be interpreted as evidence that a joint university–industry assessment has been conducted in the present study.

4.4. Indicator Weighting and Composite Evaluation Model

Once valid performance ratings have been obtained, assessment results can be aggregated hierarchically from secondary indicators to primary competency dimensions and, subsequently, to a composite evaluation score.

Indicator weights represent the relative importance of specific competencies in relation to the intended educational objectives and purposes of evaluation. They are model parameters that require appropriate justification and should not be treated as predetermined values derived from historical output counts. Because the present study did not involve expert-based weighting or individual student performance ratings, only the mathematical relationships and conditions governing model application are specified; no empirically optimal weights are assumed

Let r_{ijk} denote the valid performance rating of student j on secondary indicator i within primary dimension k, with $r_{ijk} \in \{1, 2, 3, 4\}$. Let n_k represent the number of secondary indicators within dimension k (two in the present framework), and let q_{ik} denote the weight assigned to the corresponding secondary indicator. The score R_{jk} for student j on primary dimension k is calculated as follows:

$$R_{jk} = \sum_{i=1}^{n_k} q_{ik}\, r_{ijk},\ k = 1, 2, \ldots, 6 \quad (1)$$

The weights assigned to secondary indicators must satisfy the following normalization and non-negativity constraints:

$$\sum_{i=1}^{n_k} q_{ik} = 1,\ q_{ik} \geq 0 \quad (2)$$

Let w_k denote the weight assigned to primary dimension k. The composite evaluation score S_j for student j is then defined as:

$$S_j = \sum_{k=1}^{6} w_k\, R_{jk} \quad (3)$$

The weights of the six primary dimensions must satisfy:

$$\sum_{k=1}^{6} w_k = 1,\ w_k \geq 0 \quad (4)$$

Here, R_{jk} represents student j's score on competency dimension k, while S_j represents the student's composite competency-performance score.

All weights are parameters to be determined in subsequent applications. Depending on the evaluation purpose, they may be established through expert consultation, analysis of applicable program requirements, and weight-sensitivity analysis. If the Analytic Hierarchy Process (AHP) is adopted, the composition of the expert panel, pairwise comparison matrices, and consistency-test results should be reported. In the absence of empirical weighting procedures, arbitrarily assigned values must not be presented as validated optimal weights.

Provided that all required indicators have valid ratings and that the weights are non-negative and normalized, both R_{jk} and S_j have a theoretical range of 1 to 4. For supplementary presentation, the composite score may be linearly transformed into a percentage-based scale:

$$C_j = \frac{S_j - 1}{3} \times 100\% \quad (5)$$

This transformation maps the numerical range of 1–4 onto a relative scale of 0–100% for reporting purposes only. It should not be interpreted as an empirically established competency attainment rate.

The four-level rubric is fundamentally ordinal. Calculating weighted averages of ordinal ratings therefore entails an operational assumption that differences between adjacent levels can be treated as sufficiently comparable for aggregation. Consequently, the composite score should be used as a supplementary descriptive measure rather than as a substitute for dimension-specific performance ratings or judgments concerning the attainment of critical competencies.

Where required indicators lack sufficient evidence, they should be designated as Not Assessable Due to Insufficient Evidence, rather than assigned a score of zero or excluded through weight renormalization that conceals missing information. Compliance with fundamental requirements concerning academic integrity, engineering safety, and professional ethics must be assessed independently and must not be offset by higher scores in other dimensions.

In practical applications, expert review, inter-rater reliability analysis, and weight-sensitivity testing are needed to examine the stability and interpretability of the evaluation results.

5. Evaluation Implementation Procedure

Implementation of the proposed evaluation framework involves five sequential steps.

First, the relevant educational objectives, competency indicators, and performance criteria should be established in accordance with the applicable graduate training requirements.

Second, evidence should be collected from authentic engineering tasks, including technical design documents, implementation records, validation results, and documentation of individual contributions.

Third, the collected evidence should be examined for source authenticity, technical relevance, duplication across output formats, and attribution to individual students.

Fourth, assessors with appropriate professional expertise and access to relevant evidence should independently evaluate competency performance using the common

rubric. Substantial differences in their judgments should be resolved through evidence-based review.

Finally, assessment results should be reported in a differentiated manner, including competency-dimension performance, evidence gaps, critical indicators, and compliance with non-compensatory ethical and safety requirements.

The implementation procedure follows a fundamental principle: the sufficiency of evidence must be established before a competency rating is assigned. This evidence-first approach is intended to reduce the risk of producing apparently precise numerical evaluations from incomplete documentation.

6. Evidence-to-Competency Mapping in Representative Engineering Cases

6.1. Case Materials and Relationships among Engineering Tasks

This study examined nine engineering research and development (R&D) cases documented in a 2021 case repository, supplemented by historical statistics describing relationships among different engineering outputs. According to the relevant report, 48 of the 67 master's theses completed between 2015 and 2020 were associated with graduate innovation fund support (71.6%), while 23 were additionally associated with relevant competition awards (34.3% of the 67 theses). These figures indicate potential relationships among research projects, competition entries, and master's theses, but do not directly reflect students' competency attainment. Furthermore, the aggregated historical statistics and the nine engineering cases do not constitute a unified individual-level dataset. Accordingly, they were not combined to generate individual competency scores or infer overall educational effects.

The nine cases were examined at the engineering-task level by comparing their technical objects, core engineering problems, and substantive technical content. The analysis identified six cases in which research projects, competition entries, and theses shared identifiable core technical themes or related subtasks. In the remaining three cases, the outputs involved different technical tasks or lacked sufficient evidence of a direct technical relationship (Table 3). This six-versus-three classification is descriptive and applies only to the purposively selected cases; it should not be interpreted as a population-level proportion.

Table 3. Technical Relationships among Engineering Outputs Across the Nine Cases

Case	Primary R&D project	Relationships among engineering outputs	Assessment of technical relationships
1	Electromagnetic polishing machine using liquid LMA abrasives	The competition entry and master's thesis both focused on LMA abrasive polishing equipment.	Shared core technical theme
2	Bi–Sn–In alloy casting and recovery system for machine-tool foundations	The research project, competition entry, and master's thesis all concerned the same casting and recovery system.	Shared core technical theme
3	Optimization of key structures in an FDM rapid prototyping machine	The competition entry focused on an automatic leveling device, while the thesis addressed the optimization of key machine structures.	Competition entry related to a project subtask
4	PLA filament splicing machine for FDM printers	The project, competition entry, and thesis all focused on a laser-based filament splicing device.	Shared core technical theme
5	Optimization of the extrusion head for an FDM multicolor printer	The competition entry and thesis primarily concerned an FDM-based casting system, which differed from	Different technical tasks

		the technical object of the research project.	
6	Magnetic levitation vibration-damping device for an FDM printer	The project, competition entry, and thesis all focused on an active vibration-damping device.	Shared core technical theme
7	UAV-based de-icing system for high-voltage transmission lines	The project, competition entry, and thesis all focused on a UAV-assisted de-icing device.	Shared core technical theme
8	Z-axis modularly expandable FDM printing system	The project and competition entry addressed the same technical theme, whereas the thesis focused on a laser-based filament splicing printing system.	Different technical tasks
9	FDM 3D printer using bismuth–tin alloy	The project and thesis addressed the same technical theme, whereas the competition entry concerned an intelligent robotic vehicle.	Competition entry addressed a different technical theme

These findings highlight the importance of establishing task-level relationships before determining whether materials from different output formats represent duplicate records or provide additional evidence. Neither similarities in project titles nor the involvement of the same student are sufficient, on their own, to establish continuity in competency performance across outputs.

6.2. Evidence Coverage Across Evaluation Indicators and Its Limitations

An examination of the case materials indicates that the repository provides relatively direct evidence of engineering activities relevant to selected indicators within Dimensions A1–A4. Such activities include engineering calculations, structural modeling, simulation analysis, technical solution design, functional prototype development, and performance testing. Nevertheless, the documentation of a technical activity does not necessarily provide sufficient evidence to assign a performance level.

For example, a statement indicating that simulation was conducted may establish that a particular engineering method was employed, but it does not provide sufficient information to assess the appropriateness of modeling assumptions, boundary conditions, or the reliability of simulation results. Similarly, a record stating that prototype testing was completed cannot substitute for a documented testing protocol, appropriate comparison benchmarks, or original experimental results.

For Dimensions A5 and A6, team competition awards and the presence of safety requirements within an engineering project establish relevant assessment contexts but do not, by themselves, demonstrate individual contributions or the fulfillment of professional responsibilities. Additional evidence is therefore required to support such judgments (Table 4).

Table 4. Technical Relationships among Engineering Outputs Across the Nine Cases

Evaluation dimension	Main evidence identifiable in the cases	Scope of support for competency assessment
A1 Professional Knowledge and Methodological Application	Finite element analysis, mechanical structural modeling, control methods, and material property analysis	Provides evidence of the application of disciplinary methods; the level of proficiency requires examination of original analytical outputs.

A2 Engineering Problem Identification and Analysis	Technical deficiencies in equipment, engineering requirements, proposed solution approaches, and constraint analyses	Provides evidence of problem contexts and analytical activities; the depth and quality of technical justification require further examination of supporting technical documentation.
A3 Engineering Solution Design and Implementation	Three-dimensional models, structural designs, functional prototypes, and equipment development records	Provides relatively direct evidence of engineering design and implementation activities.
A4 Experimental Validation and Practical Innovation	Prototype commissioning, experimental investigations, performance analyses, and descriptions of technical improvements	Identifies certain validation activities; judgments of performance quality depend on test data and comparative technical results.
A5 Professional Communication and Organizational Collaboration	Competition team records, thesis-related outputs, and technical presentation materials	Confirms certain collaborative and communication activities; individual contributions and actual performance require independent verification.
A6 Engineering Ethics and Professionalism	Engineering contexts involving safety, reliability, and environmental concerns	Identifies contexts involving engineering responsibilities; attainment of professional conduct requirements must be supported by additional assessment records.

These findings suggest that three distinct conditions should be reported separately: the availability of candidate evidence, the identifiability of relevant competency performance, and the sufficiency of evidence for assigning a performance level.

For indicators related to individual contributions, engineering ethics, and safety compliance, evidence is particularly likely to be incomplete in historical output-oriented records. To address these gaps, task allocation documents, design revision histories, experimental logs, risk assessments, and feedback from university and industry supervisors should be collected systematically during the training process.

When documentation is insufficient, the relevant evidence gaps should be explicitly recorded rather than interpreted as evidence of inadequate competency. Identifying missing evidence and specifying the additional documentation required for defensible judgments is therefore more valuable than attempting to assign immediate scores across all competency dimensions. Such an approach strengthens the transparency and interpretability of competency assessment.

6.3. Comparative Analysis of Three Representative Cases

6.3.1. Electromagnetic Polishing Machine Using Liquid LMA Abrasives

This case addressed the engineering requirements for deburring and polishing components with complex surface geometries. The available documentation described the design of transmission, polishing, lifting, recovery, and control mechanisms, together with three-dimensional modeling, modal analysis of the machine frame, thermal analysis of the polishing chamber, functional prototype development, and process-related experiments.

These activities provide candidate evidence relevant to A12 (Application of Engineering Methods and Tools), A21 (Identification of Engineering Requirements and Constraints), A31 (Technical Solution Design), A32 (Engineering Implementation and System Integration), and A41 (Experimental Design and Technical Validation). The case illustrates how a single engineering R&D task may generate evidence relevant to multiple competency dimensions.

However, determining whether the technical solutions were appropriate, the experimental conclusions were reliable, or the individual student's contributions were substantial would require further examination of the original simulation settings, experimental data, design iteration records, and documentation of task allocation. Competition awards or patent records may provide supplementary confirmation of engineering outputs, but cannot independently justify assigning an advanced competency-performance level.

6.3.2. Magnetic Levitation Vibration-Damping Device for an FDM Printer

This case addressed vibration-related problems during the FDM printing process. The documented engineering activities included the development of an active three-dimensional vibration-damping solution, the design of mechanical structures and an STM32-based control system, prototype fabrication, and performance testing. These activities provide candidate evidence corresponding to A22 (Engineering Problem Analysis and Justification), A12 (Application of Engineering Methods and Tools), A31 (Technical Solution Design), A32 (Engineering Implementation and System Integration), and A41 (Experimental Design and Technical Validation).

The case demonstrates the need to distinguish three levels of engineering performance assessment: developing a technical solution, implementing the proposed system, and validating its technical effectiveness.

Design documentation may provide evidence of technical choices and design decisions, while prototype records may establish that engineering implementation activities were undertaken. However, judging the effectiveness of the vibration-damping system requires verifiable performance data, including vibration response measurements, appropriate comparison conditions, and printing accuracy results.

A general statement that the experimental results were satisfactory is therefore insufficient to establish the student's competency level in experimental design and technical validation.

6.3.3. UAV-Based De-Icing System for High-Voltage Transmission Lines

This case involved UAV-assisted operations, vibration measurements, and a resonance-based de-icing solution for high-voltage transmission lines. The documented activities provide evidence of engineering problem analysis (A22) and technical solution design (A31), while also representing a relevant assessment context for engineering safety and social responsibility (A62).

For engineering tasks of this nature, evaluation should extend beyond the functional effectiveness of the de-icing system to encompass critical engineering constraints, including operation near energized transmission lines, system reliability, operational risks, and the boundaries of safe operation.

Nevertheless, the presence of safety-related risks in an engineering task does not establish that the student has fulfilled the corresponding professional responsibilities. Defensible judgments regarding engineering safety and social responsibility require additional evidence, such as documented risk identification, technical testing and validation, and records demonstrating compliance with applicable safety requirements.

6.4. Findings and Discussion of the Case Analysis

The overall examination of the nine engineering cases and the comparative analysis of three representative cases yielded three principal findings.

First, the technical relationships among engineering outputs vary across cases. Where multiple outputs share a common technical theme, overlapping records should be identified to avoid duplicate recognition of the same technical contribution. Where the underlying engineering tasks differ, the corresponding evidence should be examined separately.

Second, different output formats may provide distinct forms of evidence relevant to the same competency indicator. Their comparability should be established through

common behavioral expectations and quality criteria, rather than through the type, status, or external recognition of the outputs themselves.

Third, historical engineering case materials are generally more suitable for identifying the technical activities undertaken than for determining the quality of those activities, the extent of individual contributions, or compliance with fundamental professional requirements. This distinction reveals a substantive gap between evidence availability and the assessability of competency performance.

The present case analysis therefore established content-based relationships between evaluation indicators and documented engineering activities, while identifying the additional evidence required to support assessment judgments for different indicators.

Although the case materials provide preliminary support for the contextual applicability of the proposed evaluation framework, they do not provide a sufficient basis for assigning individual competency-performance levels, calculating composite evaluation scores, or drawing conclusions about the measurement validity of the model.

The principal implication is that evidence-based assessment of engineering outputs requires a traceable record of engineering tasks and individual contributions before performance ratings are assigned. Only when such records are available can common rubrics and judgments from multiple assessors be used to support defensible evaluations of students' competency performance.

7. Overall Characteristics of Institutional Development and Training Outcomes

Xijing University has established seven supporting institutional policies and arrangements concerning training programs, professional practice, and master's thesis management, including student development portfolios and collaborative assessment under the dual-supervisor system. The university has also continued to develop courses, teaching cases, and digital learning resources. These initiatives provide institutional and material support for task organization, the allocation of responsibilities, and process documentation. However, the consistency with which these arrangements are implemented requires further examination of routine supervision and quality monitoring records.

The related educational practices have also generated cumulative achievements in research innovation, technology transfer, academic publications, and disciplinary competitions (Table 3). Because some indicators cover different periods and student populations, they are used here solely to describe the scale of educational activities and the accumulation of training resources. They are not interpreted as changes in competency within a single student cohort or as effects attributable to PDCA implementation.

Table 3. Cumulative Institutional Development Indicators and Educational Outputs Associated with Training Reforms at Xijing University

Area of development	Cumulative indicators and outputs	Nature of the data and scope of interpretation
Student coverage	More than 210 graduate students were covered by the related training reforms.	Indicates the scope of educational activities across multiple years; does not represent a sample for effectiveness evaluation.
Curriculum development	More than 10 courses underwent reform; 16 teaching development projects were undertaken.	Reflects the scale of course and instructional development activities.
Teaching resources	More than 70 teaching cases, 24 sets of bilingual video-based course materials, and three instructional manuals.	Reflects the accumulation of teaching cases and digital learning resources.

Research innovation	More than 150 granted patents and 27 technology transfer or commercialization outcomes.	Represents cumulative outputs involving different periods and populations; does not establish a net improvement effect.
Academic publications and competitions	82 academic papers and 76 competition awards at the provincial level or above.	The outputs differ in type and statistical coverage and are not used to calculate growth rates.

Note: The indicators cover different periods and populations. All figures are presented as descriptive statistics and do not constitute a comparable before-and-after assessment of the training reforms.

8. Discussion and Conclusions

8.1. Key Findings and Their Implications for Educational Assessment

The central finding of this study is that outcome-based assessment should extend beyond quantitative statistics on research outputs to establish a defensible chain of inference linking training objectives to actual engineering performance and, subsequently, performance evidence to quality judgments. The proposed framework of six dimensions and twelve indicators provides a common evaluative language for diverse engineering activities. Task-level deduplication rules prevent multiple records originating from the same technical work from being counted repeatedly, while the hierarchical classification of evidence constrains the inferences that can legitimately be drawn from the existence of an output to a judgment about competency level. Compared with approaches that assign separate scores to courses, projects, or degree theses, the proposed framework has the conceptual potential to improve consistency in the units of assessment and transparency in the interpretation of evaluation results. However, these potential advantages remain to be examined through comparative studies involving actual assessment practices.

For educational practice, assessment reform should shift its emphasis from "increasing output requirements and scoring rubrics" to "improving the generation and use of competency-related evidence." On the one hand, engineering requirements analyses, design revisions, simulation conditions, raw experimental data, field records, risk assessments, and statements of individual contributions should be incorporated into process-oriented portfolios. On the other hand, the complementary evidentiary strengths of academic supervisors, industry mentors, and technical experts should be recognized within a shared assessment rubric, with review and reconciliation mechanisms established to address disagreements in evaluative judgments. Previous research on outcomes assessment has emphasized the traceability of assessment data and computational procedures [21]. Nevertheless, assessment quality cannot be ensured through computational procedures alone; it also requires appropriate evidence collection mechanisms and the informed professional judgment of assessors.

8.2. Research Limitations and Directions for Future Research

This study has three principal limitations. First, the nine cases were drawn from a single educational context and historical records, resulting in a highly selective sample that does not support population-level generalization. Second, although the proposed six dimensions and twelve indicators can be traced to theoretical foundations, training objectives, and engineering tasks, they have not yet undergone systematic expert consultation, content validity evaluation, or pilot implementation across different educational institutions. Third, the historical materials lack comprehensive individual-level primary records covering all indicators, independent quality assessments, and ratings from multiple assessors. Consequently, it was not possible to examine the discriminatory capacity of the performance-level rubrics, inter-rater reliability, or the measurement properties of the composite evaluation model. Nor can the findings establish a causal effect of any particular training model on improvements in student competencies.

Future research should proceed in stages. The first stage should involve small-scale pilot assessments using authentic student training records to examine the clarity of indicator definitions, the costs associated with evidence collection, and the reasons for cases being classified as "not yet assessable." The second stage should refine behaviorally anchored assessment rubrics through expert judgment and empirical assessment data, while examining inter-rater reliability across different assessor groups, the assessability of key indicators, and the sensitivity of evaluation results to indicator weights. The final stage should involve cross-contextual comparisons across different mechanical engineering specializations or educational institutions to investigate the relationships between assessment outcomes and actual professional performance in engineering practice. Where evidence is missing, priority should be given to improving the mechanisms through which relevant documentation is generated and retained, rather than producing superficially complete composite scores by assigning zero values or arbitrarily adjusting indicator weights.

8.3. Conclusions

Grounded in the principles of outcome-based education, this study developed a training quality evaluation framework comprising six competency dimensions and twelve indicators for professional master's students in Mechanical Engineering. It proposed an evidence classification approach that distinguishes output confirmation, identification of competency-related performance, and quality verification, together with rules for deduplicating evidence across different outputs based on core technical tasks, a four-level behaviorally anchored rating scheme, and a multi-assessor evaluation mechanism.

Analysis of nine historical engineering cases indicates that activities such as engineering problem analysis, solution design, system implementation, and experimental validation can yield potentially relevant evidence from different forms of outputs. However, judgments concerning technical quality, individual contributions, and fulfillment of ethical and safety responsibilities generally require supplementary primary documentation.

The study establishes an evaluation approach characterized by "common competency standards, differentiated evidence from diverse outputs, and quality judgments within clearly defined evidentiary boundaries." This approach may provide a useful reference for shifting the evaluation of postgraduate training quality from output-based statistics toward competency-oriented evidence analysis. Nevertheless, the effectiveness of the proposed framework in actual scoring practices and its generalizability across educational contexts remain to be validated through further empirical research.

Funding: This research was supported by the General Research Project of the National Alliance for Graduate Education Development of Application-Oriented Universities (Grant No. AGED2024YB45) and the Graduate Education and Teaching Reform Research Project of Xijing University (Grant No. 2023-YJG-01).

References

1. Ministry of Education of the People's Republic of China, “Opinions on Further Promoting the Classified Development of Graduate Education for Academic and Professional Degrees,” Doc. No. Jiaoyan [2023] 2, 2023. [Online]. Available: https://www.moe.gov.cn/
2. Office of the Academic Degrees Committee of the State Council, “Opinions on Effectively Organizing Degree Applications by Professional Degree Postgraduates Based on Practical Achievements,” Doc. No. Xueweiban [2026] 26, 2026. [Online]. Available: https://www.moe.gov.cn/srcsite/A22/moe_826/202610/t20261019_1455041.html
3. National Steering Committee for Engineering Professional Degree Graduate Education, “Basic Requirements for Master's Theses and Practice-Based Achievements Submitted for Professional Degree Applications in Engineering (Trial),” Doc. No. Gongcheng Jiaozhiwei [2026] 10, 2026. [Online]. Available: https://meng.tsinghua.edu.cn/xxfb/tzgg/6347.htm
4. M. Mahrishi, S. Ramakrishna, S. Hosseini, et al., “A systematic literature review of the global trends of outcome-based education (OBE) in higher education with an SDG perspective related to engineering education,” Discover Sustainability, vol. 6, art. no. 620, 2025. doi: 10.1007/s43621-025-01496-z.

5. J. G. Halls, C. Tomás, J. S. Owen, and K. Hawwash, "Mapping out the landscape of literature on assessment in engineering education," European Journal of Engineering Education, vol. 47, no. 3, pp. 373–393, 2022. doi: 10.1080/03043797.2021.2009775.
6. R. He and L. Wang, "Development of an evaluation index system for the professional practice competence of postgraduates pursuing professional master's degrees in engineering," Journal of Hebei University of Science and Technology (Social Sciences), vol. 25, no. 5, pp. 79–89, 2025.
7. C. Pacher, M. Woschank, B. M. Zunk, and E. Gruber, "Engineering education 5.0: A systematic literature review on competence-based education in the industrial engineering and management discipline," Production & Manufacturing Research, vol. 12, no. 1, art. no. 2337224, 2024. doi: 10.1080/21693277.2024.2337224.
8. M. D. Koretsky, C. J. McColley, J. L. Gugel, and T. W. Ekstedt, "Aligning classroom assessment with engineering practice: A design-based research study of a two-stage exam with authentic assessment," Journal of Engineering Education, vol. 111, no. 1, pp. 185–213, 2022. doi: 10.1002/jee.20436.
9. L. Y. Y. Luk and C. K. Y. Chan, "Assessing student learning in work-integrated learning (WIL): A systematic review on assessment methods and purposes," Studies in Educational Evaluation, vol. 81, art. no. 101345, 2024. doi: 10.1016/j.stueduc.2024.101345.
10. D. Vlachopoulos and A. Makri, "A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations," Studies in Educational Evaluation, vol. 83, art. no. 101425, 2024. doi: 10.1016/j.stueduc.2024.101425.
11. S. Deo and K. Hölttä-Otto, "Critical thinking assessment in engineering education: A Scopus-based literature review," Journal of Mechanical Design, vol. 146, no. 7, art. no. 072301, 2024. doi: 10.1115/1.4064275.
12. W. Hwang, P. Aguinalde, K. J. Crippen, et al., "Systematic review of work-integrated learning in undergraduate engineering education," Journal of Engineering Education, vol. 115, no. 1, art. no. e70050, 2026. doi: 10.1002/jee.70050.
13. J. Biggs, "Enhancing teaching through constructive alignment," Higher Education, vol. 32, pp. 347–364, 1996. doi: 10.1007/BF00138871.
14. Worldwide CDIO Initiative, "CDIO Syllabus 3.0," 2022. [Online]. Available: https://cdio.org/content/cdio-syllabus-30
15. M. T. Kane, "Validating the interpretations and uses of test scores," Journal of Educational Measurement, vol. 50, no. 1, pp. 1–73, 2013. doi: 10.1111/jedm.12000.
16. Xijing University, "Training Program for Postgraduates Pursuing a Professional Master's Degree in Mechanical Engineering (2023 Edition)," 2023.
17. H. F. Hsieh and S. E. Shannon, "Three approaches to qualitative content analysis," Qualitative Health Research, vol. 15, no. 9, pp. 1277–1288, 2005. doi: 10.1177/1049732305276687.
18. G. A. Bowen, "Document analysis as a qualitative research method," Qualitative Research Journal, vol. 9, no. 2, pp. 27–40, 2009. doi: 10.3316/QRJ0902027.
19. A. Hu, Q. Liu, and B. Daniel, "Digital technologies in authentic assessment in higher education: A systematic literature review and narrative synthesis," SAGE Open, vol. 15, no. 3, 2025. doi: 10.1177/21582440251357198.
20. D. K. Biswal, B. R. Moharana, and K. Muduli, "Development of a framework for assessing the degree of course and program outcome attainment utilizing outcome-based education framework," Journal of Engineering Education Transformations, vol. 38, no. 3, pp. 158–170, 2025. doi: 10.16920/jeet/2024/v38i3/24252.
21. S. K. Berkaya, "An integrated framework for automated measurement and prediction of program outcome attainment in engineering education," Computer Applications in Engineering Education, vol. 33, no. 6, art. no. e70094, 2025. doi: 10.1002/cae.70094.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.