

Construction of an Evaluation Indicator System for Graduate Supervisor Guidance Competence Based on Multidimensional Metrics

Na Zhang^{1,2,*}, Zhihong Liu^{1,2}

¹Engineering Research Center of Large Energy Storage Technology, Ministry of Education, Inner Mongolia, Hohhot, China

²College of Electricity, Inner Mongolia University of Technology, Inner Mongolia, Hohhot, China

*Corresponding Author

Abstract: As the key figure in graduate student training, a grad advisor's guidance skills directly affect the quality of student training, the level of discipline development, and the university's core competitiveness in talent cultivation. Currently, evaluation of some domestic university supervisors has limitations such as a single evaluation subject, an emphasis on research output over the guidance process, and difficulty balancing individual assessment with macro-level team evaluation. This article combines perspectives from peer professional evaluations, student learning experiences, and actual graduate training outcomes to build a multi-dimensional evaluation system. This system balances academic rigor with consideration of the mentoring experience and students' comprehensive development, while simultaneously achieving the dual objectives of micro-level assessment of individual advisors and macro-level evaluation of the university's advisor cohort. The scientific validity and practical applicability of this system are examined through empirical research.

Keywords: Graduate Supervisor; Mentoring Competence; Multidimensional Indicators; Evaluation System; Peer Review

1. Introduction

As China's postgraduate education enters a phase of high-quality development, strengthening the management of supervisor positions and improving the evaluation of supervisors' mentoring capabilities has become a key focus in the reform of degree-granting and postgraduate education [1]. The Ministry of Education has explicitly called for reinforcing supervisors' responsibilities in fostering moral integrity and

cultivating talent, as well as establishing a diversified and comprehensive assessment mechanism for supervisors [2].

The evaluation of graduate student supervisors abroad generally falls into two categories: a 'control-oriented' approach and a 'development-oriented' approach. In the United States, the predominant model is problem-driven developmental assessment, which relies on teaching development centers to provide diagnostic feedback; however, the evaluation results are not mandatory for supervisor qualifications. In the United Kingdom, a mandatory admission-based evaluation system is implemented, linking training assessments with supervisor professional qualifications [3]. In Australia, a periodic qualification registration review process is adopted, alongside the implementation of tiered and categorized evaluations. Canada relies on the academic community to carry out consultative and gradual evaluations. But many countries generally face issues like struggling to distinguish between a mentor's research ability and actual guidance skills, mandatory evaluations becoming too formal, and voluntary models not being widely adopted.

Existing domestic research has focused a lot on the dimensions and methods of evaluating mentors [4]. The literature commonly includes moral integrity [5], academic level, and training quality in the evaluation framework [6] [7]. But most studies have shortcomings like relying on a single evaluator, having unclear weight between peer and student evaluations, and lacking empirical testing. Some schools' current systems focus on things like graduate thesis proposals, blind reviews, and defenses, as well as the advisor's own research performance. But they are weak in setting up process-related indicators like guidance communication, personalized

teaching, and career psychological support. Such single-dimensional evaluation approaches are prone to introducing bias in assessment outcomes. Although some universities have begun implementing dual-evaluation models involving both peers and students, they have yet to establish comprehensive, multi-dimensional indicator frameworks that encompass processes, outputs, resource support, and holistic educational outcomes. In practice, many universities rely excessively on output-based metrics - such as research publications or dissertation outcomes - for supervisor assessments; these evaluations may either depend too heavily on peer academic judgments or rely solely on student ratings. It's easy to focus too much on research achievements and overlook the teaching process, and to emphasize final results while ignoring everyday guidance, making it hard to fully capture a mentor's true guiding ability. Furthermore, current evaluations mostly focus on individual mentors and lack a broader assessment of the overall level of the school's mentor team, which isn't very helpful for universities to carry out targeted mentor training and team optimization [8].

To solve the problems mentioned above, this paper focuses on the fundamental goal of fostering virtue through education, combining perspectives from peer professional evaluations, student experience assessments, and training outcome evaluations. Following principles of comprehensiveness, objectivity, practicality, and dual adaptability, it builds a multi-dimensional evaluation index system for graduate supervisors' guidance ability, covering areas like teacher ethics, process guidance, academic research, resource support, training effectiveness, and personal development. It uses the Analytic Hierarchy Process and expert consultation to determine indicator weights, divide into four evaluation levels, and design a standardized implementation process. Conduct empirical testing with the electrical discipline of a certain university as the object, using questionnaire surveys, SPSS statistics, and reliability and validity tests for data analysis [9]. The results show that this multidimensional evaluation system has good reliability and validity. Peer reviews and student evaluations are significantly positively correlated, but each focuses on different aspects, and using multiple indicators can make up for the bias of a single evaluation. The overall mentoring ability of sample

university instructors is in a good range, but there are weaknesses in career planning guidance and influence among peers in their field.

2. Construction of a Multidimensional Evaluation Indicator System

2.1 Construction Principles

The evaluation of graduate student supervisors' mentoring capabilities (hereinafter referred to as the 'Supervisor Evaluation') shall be conducted in accordance with relevant laws and regulations, including the Education Law of the People's Republic of China, the Higher Education Law of the People's Republic of China, the Teachers Law of the People's Republic of China, the 'Opinions on Deepening the Reform of Graduate Education' issued by the Ministry of Education, the National Development and Reform Commission, and the Ministry of Finance, as well as the 'Opinions on Strengthening the Construction of the Degree and Graduate Education Quality Assurance and Supervision System' issued by the Academic Degrees Committee of the Ministry of Education [10]. To ensure the scientific validity, rationality, and operability of this evaluation system, and in light of the core essence of graduate student supervisors' mentoring capabilities as well as the actual training practices at higher education institutions, the following four construction principles shall be adhered to:

1. Comprehensive Principle: This principle takes into account the ethical baseline for faculty members, the mentoring process, academic research, resource availability, program outcomes, and the holistic development of students; it integrates multi-source evidence from peers, students, and training records to encompass the full spectrum of a supervisor's mentoring capabilities.

2. Objectivity Principle: Indicators should be as observable and measurable as possible; anonymous evaluation and conflict-of-interest avoidance mechanisms shall be implemented to minimize subjective bias arising from personal favoritism in scoring.

3. Practicality principle: The indicators are expressed clearly and concisely, and the questionnaires and score sheets are easy to implement, making them convenient for regular use in university assessments.

4. Dual Adaptation Principle: It can be used both

to assess and grade individual mentors, and to aggregate all sample data to analyze the overall level of mentors across the school and various disciplines.

2.2. Evaluation Subject and Weighting Scheme Design

The multidimensional system sets up three types of evaluation information sources: peers and mentors in the same field with no conflicts of interest, the graduate students being supervised (currently enrolled or graduated within the past three years), and objective training record data. Peer evaluations emphasize academic scrutiny, focusing on assessing the concept of teacher ethics, the standardization of guidance, academic research level, and conditions of research resources. Student evaluations focus on the learning guidance experience, looking at communication feedback, personalized teaching, humanistic care, and career-psychology support. The objective archive indicators rely on the college's academic data, collecting actual

training outcomes of graduate students such as mid-term, blind review, graduation, awards, and employment. By integrating expert consultation with the Analytic Hierarchy Process (AHP), the formula for calculating the comprehensive score is determined as follows:

$$T = P \times 30\% + S \times 30\% + O \times 40\% \quad (1)$$

Where, T means total score. P means peer review score. S means student evaluation score. O means objective learning outcomes indicator score.

2.3 Specific Indicator Construction

Peer Review Subsystem (focuses on objective professional dimensions; 6 dimensions; total score: 100 points), as shown in Table 1. Student Evaluation Subsystem (Focusing on the Subjective Experience Dimension; 7 dimensions; Total score: 100 points), as shown in Table 2. Graduate Education Outcomes Assessment Subsystem (4 dimensions; total score: 100 points), as shown in Table 3.

Table 1. Peer Review Subsystem

	Level 1 Indicator	value	Explanation of Indicator Definitions
1	Professional Ethics and Teaching Conduct versus Educational Philosophy	15 points	Adherence to professional and academic ethics, a strong sense of responsibility, harmonious teacher–student relationships, equitable treatment of students
2	Guidance System and Process Standardization	15 points	Strictly implementing quality control at all key stages of the training process, ensuring adequate guidance frequency, conducting research integrity education, and improving the management of graduate student training records.
3	Academic Qualifications and Research Supervision Capacity	20 points	Staying abreast of disciplinary frontiers; guiding project design; training in research logic; evaluating the innovative aspects of research papers; and providing professional academic guidance.
4	Research Conditions and Resource Support	20 points	Funding allocation for research project platforms; provision of academic exchange opportunities; organization meetings and team structure development; and support for research-related software and hardware resources.
5	Teaching and Curriculum Contributions	15 points	Curriculum quality, teaching reform and textbook development, integrating research findings into teaching practice, and participation in revising training programs.
6	Peer Recognition and Comprehensive Evaluation	15 points	Peer recognition based on internal disciplinary assessments, support provided to junior mentors, participation in disciplinary development initiatives, and the overall effectiveness of holistic education.

Table 2. Student Evaluation Subsystem

	Level 1 Indicator	value	Explanation of Indicator Definitions
1	Professional ethics and teaching conduct versus humanistic care	15 points	Respect students' dignity; prioritize their physical and mental well-being; ensure fair allocation of honors and research support opportunities; respect students' aspirations for personal development.
2	Guidance effectiveness and communication feedback	20 points	Guaranteed frequency of guidance, patient and timely communication, actionable guidance and recommendations, and proactive follow-up on research progress
3	Academic Guidance and Competency Development	20 points	Providing literature review and research topic guidance; fostering independent research thinking; offering comprehensive revision

			support throughout the thesis writing process; and imparting academic writing skills and formatting standards.
4	Personalized guidance and tailored approaches	15 points	Tailor personalized guidance plans based on students' academic foundations and career aspirations, address individual learning gaps, respect students' research choices, and foster their ability to make independent decisions.
5	Academic Integrity and Normative Guidance	10 points	Academic integrity education, addressing research misconduct, and safeguarding students' rights to proper attribution of their research contributions.
6	Career Planning and Career Guidance	10 points	Provide advice and resources for employment, pursuing a doctoral degree, or studying abroad; guide job search skills; and address students' practical needs for career counseling.
7	Research Infrastructure and Rights Protection	10 points	Secure research resources, allocate research tasks rationally, disburse research assistant subsidies on time, foster a positive team environment, and address students' concerns regarding their treatment and compensation.

Table 3. Graduate Education Outcomes Assessment Subsystem

	Level 1 Indicator	value	Explanation of Indicator Definitions
1	Graduate Student Research Progress and Graduation Management	25 points	Monitor the on-time graduation rate of graduate students under supervision, address cases of delayed graduation, and exercise reasonable control over the research progress of graduate students.
2	Effectiveness of thesis quality control mechanisms	25 points	Thesis quality, overall performance of supervised thesis activities – including topic proposal submission, mid-term review, blind peer review, and defense; thesis sampling pass rate.
3	Output of Graduate Student Research Competitions	25 points	Publication of academic papers, patent grants, and discipline-specific innovation and entrepreneurship competitions.
4	post-graduation development	25 points	Career development outcomes for postgraduate graduates; the proportion of graduates pursuing further academic studies; the share of graduates securing high-quality employment in key industries.

2.4. Evaluation Levels and Implementation Process

2.4.1 Evaluation levels

Evaluation Grading Scale: The evaluation scale is divided into four levels: Excellent (90–100), Good (80–89), Pass (60–79), and Fail (below 60). The criteria for triggering a 'Fail' grade refer to the standard practices adopted by domestic higher education institutions and include scenarios such as professional ethics violations, academic misconduct, failure in random thesis inspections, or major incidents during the student training process – all of which constitute 'one-vote veto' situations.

Individual Evaluation: The final comprehensive evaluation score for each individual supervisor corresponds to one of the four aforementioned levels and serves as a key basis for supervisor performance appraisal, training, and promotion;

Overall Evaluation: This involves aggregating the final comprehensive evaluation scores of all supervisors within a given university, calculating metrics such as the average score, standard deviation, excellence rate, and passrate, thereby

reflecting the overall supervisory competency level of the university's supervisors and serving as an important reference for the university in optimizing its supervisor development program.

2.4.2 Evaluation implementation process

(1) **Preparation Phase:** Establish an evaluation working group, select peer experts who have no conflicts of interest, design standardized questionnaires, and conduct training sessions for the evaluation personnel.

(2) **Implementation Phase:** Anonymous peer review via questionnaires; online anonymous evaluations conducted by current and previous graduate students; the college extracts objective educational outcome data to assign scores to objective indicators; invalid questionnaires are excluded.

(3) **Data processing:** Calculate each mentor's overall score and grade; summarize all mentor data to get average scores, excellence rates, and other overall indicators.

(4) **Feedback and closed-loop improvement:** Give individual feedback to mentors about their score weaknesses. Provide overall analysis reports to the Graduate School and the college. Link the

evaluation results to mentors' admission qualifications, awards, and performance.

2.4.3 Explanation of the dual adaptability of the evaluation system

This evaluation system uses the logic of 'individual evaluation - aggregated analysis - overall assessment' to achieve a dual evaluation of both individual mentors and the overall guiding ability of universities. The steps are as follows:

(1) Individual evaluator assessment

For individual mentors, it use multi-dimensional evaluations to calculate their final overall score and rating, fully reflecting their mentoring abilities, and providing an objective basis for their assessment, training, and promotion.

(2) Overall assessment

For each subject, gather evaluation data from all individual advisors, calculate metrics like the average score, excellence rate, pass rate, and standard deviation for advisors' guidance ability, analyze the overall level, strengths, and weaknesses of the advisors, as well as the differences in guidance ability among advisors with different titles, to provide data support for the university's overall management.

This dual-adaptability design not only solves the problem in the current evaluation system of 'focusing on individuals while neglecting the overall picture,' but also combines individual evaluation with overall evaluation, meeting both the micro-assessment and macro-management needs of universities.

3. Empirical Research

3.1 Research Subjects and Instruments

Select the electrical engineering department of a local university. The survey targets include on-duty graduate advisors, peer advisors in the same field, and graduate students. 30 peer questionnaires were distributed, and 27 valid responses were collected. Additionally, 110 student questionnaires were distributed, all of which yielded valid responses (110 responses). The research instruments consisted of a self-developed multidimensional evaluation questionnaire for peers and a student questionnaire, both utilizing a five-point Likert scale; these instruments underwent expert review and preliminary pilot testing for optimization. Statistical analyses—including reliability and validity assessments, descriptive statistics, and correlation analysis—were performed using

SPSS 26.0.

3.2 Empirical Data Analysis

3.2.1 Reliability test

For the overall evaluation system, Cronbach's α coefficient is 0.891; for the peer evaluation sub-system, $\alpha = 0.864$; and for the student evaluation sub-system, $\alpha = 0.878$. The α coefficients for all first-level dimensions exceed 0.8, indicating good internal consistency of the questionnaire.

3.2.2 Validity tests

The KMO measure for overall suitability is 0.861, and the Bartlett's test of sphericity yields a p-value < 0.001 , indicating that factor analysis is appropriate; the extracted common factors align well with the framework of the six primary indicators, with a cumulative variance contribution rate of 79.12%; all indicator factor loadings are > 0.7 ; both convergent validity and discriminant validity meet the required standards; and the indicator system is structurally sound.

3.2.3 Descriptive statistic

The average comprehensive evaluation score for university supervisors was 81.72; the excellence rate was 15.74%, the good rate was 58.33%, and the passing rate was 25.93%; no samples failed to meet the criteria, indicating that supervisors as a whole are at an 'good' level. By dimension, the 'Professional Ethics and Conduct' dimension scored the highest, while the 'Career Planning and Employment Guidance' and 'Academic Peer Recognition' dimensions scored relatively low.

3.2.4 Correlation analysis

The Pearson correlation coefficient between peer evaluation scores and student evaluation scores is 0.674 ($p < 0.001$), showing a significant positive correlation, but there are still differences between them. Peers emphasis on the mentor's academic level and research resources more. Students are paying more attention to communication experiences, personalized guidance, and career support. This shows that multi-dimensional and multi-source evaluation can achieve complementary perspectives and reduce the bias that comes from a single evaluator. Objective output indicators are significantly related to academic guidance, which shows that things like papers and jobs are inherently connected to the quality of everyday guidance.

3.3 Empirical Summary

Based on the above empirical data analysis, the following conclusions are drawn,

1. The graduate advisor guidance ability evaluation system built in this study, based on multiple indicators, has good reliability and validity. The Cronbach's α coefficients are all above 0.86, and the cumulative variance contribution rate of the common factors is over 79%, indicating that the evaluation system is scientific, reasonable, and stable, and can objectively and accurately reflect the guidance ability of graduate advisors.

2. Peer evaluations and student evaluations are significantly positively correlated, with each having its own focus and complementing each other. Peer evaluations focus on academic professionalism and objective aspects, while student evaluations emphasize the guidance experience and subjective aspects. Objective training outcomes are significantly related to the academic guidance dimension, indicating that outputs like papers and employment are inherently linked to the quality of daily guidance. Multidimensional evaluations can comprehensively cover the core aspects of a mentor's guidance ability, going beyond the limitations of evaluations from a single perspective.

3. The evaluation system can provide dual assessments of both individual mentors and the overall guidance ability of a university's disciplines. Individual evaluations can accurately reflect a mentor's strengths and weaknesses, while overall evaluations give a complete view of the discipline mentors' guidance capabilities.

4. The survey results show that the graduate supervisors at the school generally have good mentoring abilities, but there are still shortcomings in areas like peer recognition and career planning guidance. This indicates that the evaluation system can accurately identify problems in mentoring and provide directions for improvement, further confirming the system's practicality and effectiveness.

4. Recommendations

1. Dynamically optimize multi-dimensional indicators to reflect disciplinary differences. This study uses weights from the general recommendation version, and different disciplines can tweak each dimension's weight based on their characteristics; refine some process-related indicators (like specifying the minimum supervision frequency for master's and doctoral students); set up a mechanism for dynamically revising indicators to adapt to new

requirements in graduate education reform.

2. Strictly standardize the evaluation implementation process. Implement anonymous and peer avoidance systems, increase the sample size of peer experts; conduct pre-evaluation training to reduce misunderstandings; establish rules for handling invalid questionnaires and extreme abnormal scores to ensure the evaluation is objective and fair.

3. Strengthen the closed-loop application of evaluation results. Provide individual feedback to mentors regarding their strengths and weaknesses; offer targeted mentoring skills training for mentors with relatively low scores; utilize multi-dimensional evaluation results as a key basis for determining mentors' eligibility for student recruitment, award nominations, and position evaluations, thereby shifting away from an 'exclusively focused on research outputs' approach; and leverage collective team evaluation data to provide a decision-making basis for the university's mentor recruitment and training planning efforts.

4. Addressing gaps in mentors' educational capabilities: In response to common issues identified through empirical research — such as inadequate career planning guidance — conduct specialized mentor training sessions; guide mentors to balance their research responsibilities with student psychological counseling and employment support, thereby fulfilling the comprehensive requirements of fostering virtue and cultivating talents.

5. Conclusion and Future Directions

This study has developed a multidimensional indicator system for assessing mentor guidance capabilities that integrates peer review, student-led learning experiences, and objective outcome-based evaluations, thereby serving the dual purpose of evaluating individual mentors as well as conducting an overall assessment of the university's mentorship team. Empirical results demonstrate that this system exhibits good reliability and validity; moreover, its multi-dimensional, multi-source indicators effectively mitigate the limitations inherent in single-dimensional evaluation approaches.

This empirical study focused solely on a single local university; future research could expand the sample to include universities of different types; further exploration could be conducted on how big data can assist in process-oriented

evaluation; deeper investigation could be carried out into the linkage mechanism between evaluation results and mentor incentives and training programs; and efforts should continue to refine the evaluation system for graduate student supervisors' guidance capabilities.

Acknowledgments

This work is supported by Inner Mongolia Autonomous Region Postgraduate Education and Teaching Reform Project (No.JG2024032C).

References

- [1] Sun Mingming, Yang Haoran. A Study on Strategies for Enhancing the Supervision Competence of Young Graduate Supervisors in Local Higher Education Institutions. *Journal of Heilongjiang Institute of Teacher Development*, 2025, 44(6): 14-18.
- [2] Li Yongfeng, Wang Zhankui, Li Bo. The Construction of an Innovative Postgraduate Supervisor Team in the Context of Building China into a Leading Country in Education. *Journal of Henan Institute of Science and Technology*, 2025, 45(10):32-40.
- [3] Song Xinguo, Ye Shuping, Zhang Chuanxiang, Experiences and Insights from the British and American Graduate Supervision Systems. *Journal of Huainan Vocational Technical College*, 2017, 17(1):108-110.
- [4] Yang Fei, Shi Beibei. Construction of an index system for evaluating supervisors' performance of responsibilities from the perspective of graduate students. *Journal of Science and Education*, 2025, (15):30-34.
- [5] Feng Jingyi. Exploring the pathways for enhancing the educational competence of early-career graduate student supervisors from the perspective of competency models. *Journal of Liaoning Normal University (Natural Science Edition)*, 2026, 49(2):275-282.
- [6] Li Bin, Xie Shuibo, Jiang Shuyuan. Problems and Countermeasures of the Postgraduate Tutors' Appraisal System in the Universities From the Perspective of Dedication of Moral Education. *Education Modernization*, 2019, 6(63):149-153.
- [7] Ma Huanling, Huang Lijing. The Construction of the Evaluation Index System for the Postgraduate Tutors' Duty of Morality-based Education. *Modern Education Management*, 2020(8):84-92.
- [8] Wang Fayuan. An Analysis of the Construction of an Assessment and Evaluation Indicator System for Master's Degree Supervisors in Higher Education Institutions. *Journal of Higher Education*, 2016, (3):4-6.
- [9] Yang Chenggen, Yang Qin. Application of SPSS Project Analysis in Questionnaire Design. *Journal of Higsher Correspondence Education (Natural Sciences)*, 2010, 23(3): 107-109.
- [10] General Office of the Ministry of Education. Notice on Further Standardizing and Strengthening the Management of Graduate Student Training. [2019-2-26].