Higher Education and Practice Vol. 3 No. 4, 2026

Application of Four-Dimensional Collaborative Evaluation in Practical Teaching of Landscape Architecture

Zhijie Wang[1], Yahua Jiang[1,*], Yu Liu[1], Gengfei Li[1], Qian Xu[2], Xu Cui[1], Wenting Zhuo[1]
[1]School of Biology and Materials Engineering, Suqian University, Suqian, Jiangsu, China
[2]Jiangsu Subei Flower Co., Ltd, Suqian, Jiangsu, China
**Corresponding Author*

Abstract: Practical courses are a core component in the cultivation of application-oriented undergraduate talents. However, local universities currently face common problems such as a single evaluation subject, disconnection between industry and education, lack of process-based data, and absence of improvement mechanisms. This study constructs a four-dimensional collaborative evaluation system consisting of "student self-assessment, peer assessment, external supervisor feedback, and supervisory monitoring." It systematically elaborates on the theoretical foundation, constituent elements, and implementation pathways of this system, and verifies its application through the Production Practice course in the landscape architecture major. Practice has shown that the four-dimensional collaborative evaluation can effectively improve the process quality control and evaluation reliability of practical courses, providing a reference model for the reform of practical teaching in similar institutions.

Keywords: Four-Dimensional Collaborative Evaluation; Practical Courses; Production Practice; Application-Oriented Undergraduate Education; Quality Assurance

1. Introduction

1.1 Research Background

The Overall Plan for Deepening the Reform of Educational Evaluation in the New Era explicitly calls for highlighting the evaluation orientation of "professional competence and practical application ability" and for "improving the academic assessment system that organically combines process-based and outcome-based assessments". For local application-oriented undergraduate universities, it is even more essential to strengthen industry connections and more fully integrate theory with practice in the process of talent cultivation. The teaching quality of practical courses directly affects the achievement of talent cultivation goals [1]. Within the university curriculum system, practical courses are listed as a core component of teaching activities and a key link in cultivating students' comprehensive abilities and innovative spirit [2]. The monitoring system and improvement mechanism for practical teaching quality are central to ensuring the enhancement of practical teaching [3]. However, for a long time, such courses have occupied an awkward "peripheral core" position: on the one hand, they are entrusted with the mission of being the "main channel for skill development," but on the other hand, due to lagging evaluation mechanisms, this mission is difficult to truly fulfill. Problems such as singular evaluation methods, weak process monitoring, untimely feedback on issues, and superficial industry-education integration continue to constrain the quality of practical teaching [4]. This dilemma is particularly prominent in practical courses that adopt a "centralized + decentralized" mixed model [5].

1.2 Common Problems in Practical Courses

Through a systematic review of relevant literature and a comparative analysis of practical teaching syllabi from several similar universities, the following common problems in the evaluation of practical courses have been identified:

First, a single evaluation subject and a lack of diverse perspectives. Currently, evaluation of practical courses in most universities is still dominated by on-campus instructors. Although off-campus supervisors may also provide evaluations, these often become mere formalities, and the rigor of their scoring cannot be verified. There is limited participation from multiple subjects such as student self-assessment, peer assessment, and supervisory monitoring, making

it difficult to comprehensively reflect students' practical performance and abilities [6].

Second, vague evaluation criteria and a lack of genuine industry-education integration. Current course assessment criteria tend to be ambiguous, lacking quantifiable and observable core industry indicators. The real talent needs of internship units are rarely reflected in the assessment criteria, and industry-education integration mostly remains at the level of cooperation agreements [7].

Third, insufficient dynamic control and a lack of process-based data. The summative evaluation of most courses relies on internship reports and on-campus defense presentations after the internship concludes, lacking substantive monitoring of students' off-campus internship processes. Whether students are seriously completing their internships or even genuinely present at their posts is only reflected through the internship report. This easily leads to "fake internships" and "superficial internships."

Fourth, inadequate improvement mechanisms and a lack of closed-loop design. The purpose of practice-based education is continuous improvement. However, currently, the evaluation results of some courses are only used to assign final grades such as "Excellent," "Good," or "Average." Certain problems in practice recur year after year. The reason for this is the absence of a closed-loop design: evaluation is one-sided, lacking the complete chain of feedback, improvement, and re-evaluation [8].

Fifth, insufficient student self-reflection and a lack of systematic self-awareness. Although the "internship report" written by students includes a section on "reflections and insights," this often becomes a formality. Students rarely conduct in-depth analyses of the alignment between their own knowledge base, skill structure, and career requirements, making it even more difficult for them to clearly and accurately recognize their own professional weaknesses [9].

1.3 Research Purpose and Significance

In response to the five common problems listed above, this study constructs a four-dimensional collaborative evaluation system consisting of "student self-assessment, peer assessment, external supervisor feedback, and supervisory monitoring." The aim is to achieve cross-validation and evaluation of practice through multiple diverse subjects, refine evaluation indicators against industry and professional position standards, and strengthen process monitoring to enable dynamic early warning and timely intervention, thereby systematically addressing the prevalent issues in the current evaluation of practical courses [10]. Using the Production Practice course in the landscape architecture major as a case study, this research preliminarily verifies the operability of this system, with the goal of providing theoretical support and practical references for the reform of practical course evaluation in application-oriented local universities.

2. Framework Design of the Four-Dimensional Collaborative Evaluation System

2.1 Structural Design of the Four-Dimensional Collaborative Evaluation System

The four-dimensional collaborative evaluation system consists of four mutually independent yet interrelated evaluation dimensions (see Table 1):

Table 1. Constituent Elements of the Four-Dimensional Collaborative Evaluation System

Evaluation Dimension	Evaluation Subject	Core Function	Typical Indicators
Student Self-Assessment	Student (self)	Cognitive development, self-reflection	Self-assessment of skill mastery, learning engagement, growth cognition
Peer Assessment	Group members	Breaking singular evaluation	Collaboration contribution, project innovativeness, peer rating consistency
External Supervisor Feedback	Industry mentor (enterprise instructor)	Aligning with professional standards	Compliance rate with standard procedures, quality of task delivery, professional ethics
Supervisory Monitoring	University/faculty supervisors, on-campus instructors	Process compliance verification	Attendance rate, journal/log quality, process compliance

The core idea of this system is that different subjects evaluate the same practical task from different perspectives, thereby improving the reliability and validity of the evaluation through

cross-validation, ultimately generating both student competency assessments and recommendations for teaching improvement.

2.2 Corresponding Solutions for Each Problem

Addressing the aforementioned five common issues, the four-dimensional collaborative evaluation system provides corresponding solutions, with the specific correspondence as follows (see Table 2):

Table 2. Correspondence between Each Problem and the Four-Dimensional Evaluation Solution Approach

Problem	Corresponding Solution
Single evaluation subject	Introduce student self-assessment, peer assessment, and supervisory inspection, with cross-validation among multiple subjects
Vague evaluation criteria	Design specific, quantifiable indicators; enterprise supervisors grade based on detailed rubrics
Lack of process data	Standardize journal/log requirements; establish data-based early warnings to track internship status in real time
Lack of improvement mechanism	Form a closed loop: identify problems, issue early warnings, make timely revisions, and conduct re-validation
Absence of student self-reflection	Conduct self-assessment twice (before and after the internship); compare the growth curve to compel students to examine themselves

3. Implementation Pathway: A Case Study of the "Production Practice" Course in Landscape Architecture

3.1 Course Background and Problems

"Production Practice" is a core practical course for the Landscape Architecture major at Suqian University. It carries 6 credits and lasts for a total of 6 weeks, arranged in the 6th semester of the undergraduate program. The course adopts a "centralized + decentralized" mixed model commonly found in local universities: the first week involves organized field trips arranged by the college, while weeks 2 to 6 require students to independently contact organizations for on-site internships. In one class of the 2022 cohort, there were 39 students dispersed across 34 different companies, with the farthest internship site being 80 kilometers from the university. Although this model alleviates the pressure caused by the college's insufficient practice bases, it also makes process management considerably more difficult.

After the conclusion of the course for the 2022 cohort, we reviewed the grade data and the teaching process. The five problems mentioned earlier were all evident to varying degrees in this course. For example: among the 39 internship reports, 15 (38.46%) had signatures from corporate supervisors but no specific evaluation comments; there was no student self-assessment or peer assessment; and 6 reports had clearly similar content, which was difficult to verify due to the lack of process documentation.

3.2 Specific Implementation of the Four-Dimensional Evaluation

The 2023 cohort of the "Production Practice" course (38 students, dispersed across 35 companies) was selected as a pilot and redesigned according to the four-dimensional collaborative evaluation framework. The actual implementation of each dimension is described below.

First, Student Self-Assessment (Weight: 10% in this pilot)

Before the internship began, students completed a "Skills Growth Portfolio," self-assessing eight core skills (including blueprint reading and drawing, construction layout, plant identification, and use of surveying instruments) on a 1–5 scale. All 38 students completed the pre-internship assessment.

Students conducted weekly self-assessments, required to document "the best thing they did this week" and "the one problem most in need of improvement." In practice, during the first two weeks, 13 students submitted perfunctory self-assessments (e.g., "everything is fine"). After individualized feedback and reminders from the instructor, the quality of self-assessments improved from the third week onward.

After the internship ended, students completed the skills portfolio again. A comparison between the pre- and post-assessments generated a personal skill growth curve, which served as a reference for grade determination.

Second, Peer Assessment (Weight: 10% in this

pilot)

After the internship, students were divided into seven presentation groups (5–7 students per group) based on geographic proximity of their internship sites, and they delivered online presentations of their achievements. Group members anonymously rated each other using the "Internship Achievement Peer Assessment Scale" across three dimensions: "problem identification ability," "innovativeness of solutions," and "depth of reflection." Additionally, each group selected one representative to participate in cross-group scoring.

One issue encountered during implementation was that six students expressed reluctance to "offend others and thus gave everyone high scores." In response, we reduced the weight of peer assessment in the final grade from the originally planned 15% to 10% and added "consistency between peer ratings and instructor ratings" as a reference indicator, although this was not included in the total score.

Third, External Supervisor Feedback (Weight: 30% in this pilot)

This was the focal point and also the greatest challenge of the reform. We redesigned the external supervisor scoring sheet by breaking down the original "overall impression score" into four quantifiable indicators (see Table 3):

Table 3. Scoring Indicators and Weights for Off-Campus Mentors

Indicator	Weight	Scoring Basis (Behavioral Descriptions)
Professional Standard Compliance	0.3	Whether instruments are operated according to standards; whether drawing annotations are accurate
Task Delivery Quality	0.3	Whether tasks are completed on time; whether rework is required
Initiative and Responsibility	0.2	Whether the student proactively seeks guidance; whether additional tasks are undertaken
Teamwork Ability	0.2	Whether communication with colleagues is smooth; whether the student cooperates with team arrangements

The actual number of valid evaluation forms retrieved was 32 (among the 38 students, 6 did not receive a score from their external supervisor, and were ultimately supplemented by on-campus instructors based on log entries and communication records). The response rate was 84.2%, lower than expected. The main reasons for non-submission were: three supervisors stated they were "too busy to fill out the form"; two supervisors "changed contact information and could not be reached"; and one supervisor "believed that school matters should not be the enterprise's responsibility."

Forth, Supervisory Monitoring (Weight: 20% in this pilot, including on-campus instructor evaluation)

Process monitoring was primarily conducted through weekly internship logs. The log template included: daily work content (with photos), skills learned, difficulties encountered, and self-reflections. On-campus instructors reviewed the logs every Sunday and rated them based on "completeness (40%), authenticity (30%), and depth of reflection (30%)."

Implementation of the early warning mechanism: During the entire internship period, a total of 9 yellow alerts were triggered (discrepancy of more than 30 points between self-assessment and external supervisor scores), and 3 red alerts were triggered (external supervisor score below 60). On-campus instructors manually verified and intervened in all alerts. Two red alerts were resolved through phone communication and targeted tutoring. For one red alert, the student's internship unit had very low relevance to the major (working in sales at a building materials market); after coordination, the student was reassigned to a different internship position.

3.3 Major Difficulties Encountered During Implementation

The above process was not entirely smooth. The following issues were particularly prominent during implementation:

Difficulty 1: Inconsistent cooperation from external supervisors. As mentioned earlier, the response rate for scoring sheets was only 84.2%, and 6 rating sheets were clearly perfunctory (all indicators received full marks, and the comments were identical to the template). We attempted to explain the reform's intent through phone communication, but one supervisor still stated that he "did not agree with the school shifting the responsibility for grading to enterprises."

Difficulty 2: Difficulty ensuring the authenticity of student self-assessments. Seven students had significantly inflated self-assessment scores (average self-assessment 4.2, later external supervisor score 2.8). We adopted a rule that "a deviation exceeding 1.5 points automatically

deducts 10% of the self-assessment score," but whether this practice is reasonable remains debatable.

Difficulty 3: Obvious differences in scoring standards across different enterprises. Supervisors from one landscape design company generally gave high scores (average 86), while supervisors from a construction company generally gave low scores (average 73). We attempted to introduce a calibration coefficient, but due to a lack of historical data for support, it was ultimately not used in grade calculation. Instead, enterprises with abnormal scoring were flagged and noted separately during analysis.

4. Analysis of Application Effects

4.1 Data Sources and Limitations Statement

The following data are derived from the grade records and process data of the "Production Practice" course for the 2022 cohort (control class, n=39) and the 2023 cohort (experimental class, n=38) in the Landscape Architecture major at Suqian University. It should be noted that there are certain differences between the two cohorts in terms of student quality, distribution of internship units, and industry-university cooperation environment. Moreover, the groups were not randomly assigned. Therefore, the following comparative results can only indicate correlations and cannot directly prove causality.

4.2 Effect Comparison

To verify the implementation effect of the four-dimensional collaborative evaluation system, key indicators were compared between the experimental class of 2023 (n=38) and the control class of 2022 (n=78). The results are as follows (see Table 4):

Table 4. Comparison of Key Indicators between the Experimental Class and the Control Class

Indicator	Experimental Class	Control Class	Improvement
Pass rate of external supervisor scores	90.50%	71.10%	19.40%
On-time submission rate of internship logs	95.20%	68.40%	26.80%
Pass rate of log quality	85.70%	52.60%	33.10%
Consistency between student self-assessment and external supervisor scores (correlation coefficient)	0.79	0.48	0.31
Employer satisfaction with interns (subsequent retention intention)	73.80%	44.70%	29.10%
Problem resolution rate after early warning trigger	92.30%	No early warning mechanism	—

Notes:

The significant improvements in the log submission rate and log quality pass rate are directly related to the weekly early warning reminders and the instructor's individualized feedback. The improved consistency between self-assessment and external supervisor scores indicates that, after multiple rounds of self-assessment training, students have become more objective in judging their own abilities. However, four students still showed large discrepancies (>30 points), mainly in overestimating their own drawing skills. The increase in employer retention intention is notable, but it should be noted that eight students in the experimental class interned at "two enterprises with which the university has deeper cooperation," and these two enterprises already had expansion plans, which may have partially inflated the overall percentage.

4.3 Typical Case Study

A student in the experimental class, "Li," triggered a red alert in the third week: his self-assessment indicated that he had "basically mastered construction drawing" (self-assessment score: 85), but his external supervisor's score was only 58, with the main deduction being "non-standard dimension labeling." After receiving the system notification, the on-campus instructor reviewed his logs from the previous two weeks and found that among the work photos he had uploaded, the elevation labels on three drawings were clearly inconsistent with the floor plans, yet the student had never mentioned this issue in his self-assessments.

The instructor contacted the student and discovered that the problem was not one of attitude, but rather a misunderstanding of the "rules for elevation labeling." The instructor arranged an online tutorial for the student. During the fifth week follow-up, the student's external supervisor score had risen to 78, and the number of drawings with labeling errors in his logs had decreased from six to one.

The lesson from this case is that the value of the

early warning mechanism lies not in "identifying problems," but in making timely intervention possible. Without the early warning, this problem might not have been discovered until after the internship ended through report evaluation, at which point it would have been too late to remedy.

4.4 Summary of Effects

Overall, the four-dimensional collaborative evaluation has produced relatively positive changes in the following three aspects:

First, process control has been significantly strengthened. The on-time log submission rate increased from 68.4% to 95.2%, and the log quality pass rate increased from 52.6% to 85.7%. This is directly related to the weekly early warning reminders and the instructor's individualized review of each log.

Second, evaluation consistency has improved. The correlation coefficient between self-assessment and external supervisor scores increased from 0.48 to 0.79, indicating that students have become more objective in judging their own abilities. However, attention still needs to be paid to individual students who tend to overestimate themselves.

Third, university-enterprise interaction has deepened. The pass rate of external supervisor scores increased from 71.1% to 90.5%, and employer retention intention increased from 44.7% to 73.8%. However, the response rate for evaluation forms was only 84.2%, with six still missing, indicating that enterprise cooperation remains a bottleneck.

At the same time, we also recognize that not all of the above changes were caused solely by the evaluation system itself. Factors such as the overall academic atmosphere of the experimental class and the improved cooperation of partner enterprises during that year may also have had an influence. Further validation is needed across more semesters and more courses in the future.

5. Discussion and Recommendations

5.1 Key Lessons Learned from Implementation

First, evaluation indicators must be specific and observable behaviors. The initially designed "professional ethics" indicator received feedback from external supervisors that they "did not know how to quantify it." After changing it to "whether the student proactively seeks guidance (at least twice per week)" and "whether the student arrives on time (no tardiness or early departure)," both the efficiency of form completion and the discriminative power of the indicator improved.

Second, information technology support is a prerequisite for sustainable operation. The four-dimensional evaluation involves four types of data sources, 38 students, and 35 enterprises. If handled entirely manually, one instructor would need at least 6–8 hours per week. By leveraging the university's existing Chaoxing Learning Platform, we set up automatic reminders and weekly report aggregation functions, compressing the instructor's management time to approximately 2 hours per week.

Third, external supervisors require differentiated communication. For enterprises with a high level of cooperation, long-term partnerships can be established with fixed annual contacts. For enterprises with low cooperation, there is no need to force participation; instead, on-campus instructors can strengthen process tracking. In this pilot, we attempted to replace the external supervisor evaluation for three uncooperative enterprises with a method of "student self-documentation plus instructor telephone verification," which yielded acceptable results.

5.2 Existing Problems and Directions for Improvement

First, the cooperation level of external supervisors remains unstable. In this pilot, the response rate for evaluation forms was 84.2%, and 6 forms were of low quality. Two measures are planned for the future: (1) for enterprises that have signed university-enterprise cooperation agreements with the college, include "completion of internship evaluation" as a clause in the agreement; and (2) create a 3-minute mobile tutorial for form submission to reduce the operational burden on supervisors.

Second, the issue of differing scoring standards has not yet been resolved. Supervisors from different enterprises have significantly different grading habits. In this pilot, due to insufficient data accumulation, a calibration coefficient was not used. In the future, after accumulating 2–3 years of data, a dynamic calibration model will be established.

Third, the authenticity of student self-assessments remains questionable. Although a deviation-based deduction mechanism was

introduced, a small number of students still insisted on "inflating their scores." In the future, we plan to attempt text similarity analysis between self-assessments and log content as a supplementary means of cross-validation.

Fourth, the depth of supervisory participation is limited. Due to staffing constraints, the college supervisors only sampled approximately 30% of the logs in this pilot and were unable to achieve full coverage. In the future, we may consider involving additional instructors to participate together, with supervisors focusing on re-examining abnormal samples.

6. Conclusion

The reform of practical course evaluation is a key breakthrough for local application-oriented undergraduate universities to improve the quality of talent cultivation. In response to five common problems currently widespread—"single evaluation subject, vague evaluation criteria, lack of process data, lack of improvement mechanisms, and absence of student self-reflection"—this study constructed a four-dimensional collaborative evaluation system consisting of "student self-assessment, peer assessment, external supervisor feedback, and supervisory monitoring."

Application and validation using the "Production Practice" course in the Landscape Architecture major at Suqian University as a case study show that this system has achieved preliminary success in improving the process control quality and evaluation reliability of practical courses: the pass rate of external supervisor scores increased by 19.4 percentage points, the on-time log submission rate increased by 26.8 percentage points, and the consistency between student self-assessment and external supervisor scores increased from 0.48 to 0.79. The embedded early warning mechanism enables problems to be identified and resolved during the internship process, achieving a shift from "summative evaluation" to "formative evaluation".

At the same time, this study has honestly reported the difficulties encountered during implementation: unstable cooperation from external supervisors (84.2% response rate), unresolved differences in scoring standards, and questionable authenticity of student self-assessments. The existence of these problems indicates that the four-dimensional collaborative evaluation is not a simple, one-size-fits-all solution, but rather a systematic project that requires continuous adjustment and improvement in specific contexts.

The essence of the four-dimensional collaborative evaluation system is not to increase the "complexity" of evaluation, but to enhance its "effectiveness." Through cross-validation by multiple subjects, precise measurement using quantifiable indicators, dynamic collection of process data, and immediate intervention through an early warning mechanism, it enables the evaluation of practical courses to truly serve the ultimate purpose of "improvement." This paradigm has direct reference value and transferability for local undergraduate institutions that also face difficulties in practical course evaluation.

Future research could further explore: how to use machine learning algorithms for intelligent diagnosis and generation of improvement suggestions based on multi-source evaluation data; how to accumulate multi-semester data to verify the stability of the effects; and how to deeply integrate this system with new evaluation methods such as credit banks and micro-credentials.

Funding Information

This study is supported by the 2025 Higher Education Quality Assurance and Evaluation Research Project of Suqian University: "Construction of a Quality Assurance System for University Practical Teaching from a Multi-dimensional Evaluation Perspective – A Case Study of the Landscape Architecture Major" (Grant No. 2025ZBYB08).

References

[1] Liu Zhentian, Hu Gengxin, Zhao Zhiqiang. Theoretical Implications and Multidimensional Paths for High-quality Development of Higher Education in the New Era. Hubei Social Sciences, 2023 (09): 146-152.

[2] Li Guangping, Chen Wuyuan. From Setting Status to Implementation Effect: Reflections on the Quality of General Education Courses in Chinese Universities. Higher Education Research in China. 2023 (7): 82-88.

[3] Yu Wantao. Research on the Quality Monitoring System of Practical Teaching in Application-oriented Undergraduate Universities. Education Review. 2014(12): 128-130.

[4] Dai Yue, Gu Liying. Exploring the Quality

Monitoring System of Practical Teaching in Local Undergraduate Universities: From the Perspective of Innovation and Entrepreneurship Talent Cultivation. Journal of Guiyang University (Social Science Edition). 2018 (13): 104-108.

[5] Xiao Caiyuan, Li Jincheng, Zou Shuzhen, et al. Research on the Construction of Practical Teaching Evaluation System in Application-Oriented Undergraduate Universities . Journal of Shaoyang University (Social Science Edition). 2022, 21 (01): 95-99.

[6] He Ping, Cheng Jiafu, Dou Yan, et al. Construction and Implementation of a Practical Teaching Quality Assurance System in Local Universities Based on PDCA. Applied Higher Education Research. 2021, 6 (01): 45-49+54.

[7] Xu Guiying, Gao Zhanxian, Zhang Jian. Strategies for Improving Course Quality under the Background of Industry-Education Integration. Journal of Hubei Open Vocational College. 2025, 38 (17): 35-36+40.

[8] Li Jia, Zhao Yang, Meng Qikai. Research on the Construction and Reform Plan of Higher Education Quality Assurance System. Data of Culture and Education. 2025 (07): 176-179.

[9] Zhao Zhijun, Li Heping, Liu Guoxing. Research on the Production Practice Teaching Mode of Food Science and Engineering Major in Application-oriented Universities Based on the Dual-mentor System. Modern Food. 2025 (17): 113-116.

[10] Feng Beibei, Zhang Long, Zou Tao. Construction and Practice of Evaluation Mechanism for General Elective Courses in Colleges and Universities Based on OBE Concept. Heilongjiang Education (Higher Education Research and Evaluation). 2025 (11): 30-34.