DESIGN AND IMPLEMENTATION ASSURANCE OF A DYNAMIC INDICATOR SYSTEM FOR CLASSIFIED FACULTY PERFORMANCE EVALUATION IN THE POST-"BREAKING THE FIVE ONLYS" ERA

Lei Feng
Human Resources Department, Shaanxi University of Technology, Hanzhong 723001, Shaanxi, China.

Abstract: Addressing practical dilemmas such as homogenization, rigid indicators, and singular evaluation subjects in faculty evaluation under the background of "Breaking the Five Onlys," this study aims to construct a dynamic performance evaluation system adapted to the requirements of the connotative development of higher education. Adopting a paradigm combining normative analysis and empirical research, this study comprehensively utilizes literature analysis, the Delphi method, and the Analytic Hierarchy Process (AHP). It establishes differentiated evaluation indicators and dynamic weight configurations covering four types of faculty: teaching-focused, research-focused, social service-oriented, and teaching-research balanced. By constructing a "classification-dynamic-synergy" three-dimensional evaluation model, the study designs an implementation assurance mechanism encompassing institutional, organizational, technical, and cultural dimensions, and verifies the system's effectiveness through pilot applications. The results indicate that this dynamic evaluation system possesses good reliability and validity, significantly improves faculty satisfaction with evaluations, and optimizes the structure of teaching and research outputs. This study breaks through the traditional static evaluation paradigm, achieves the unification of theoretical innovation and practical application, and provides a scientific basis and actionable scheme for universities to deepen the reform of faculty evaluation systems in the new era and improve the modernization level of governance capabilities.
Keywords: Breaking the five onlys; Faculty performance evaluation; Classified assessment; Dynamic indicator system; Implementation mechanism

1 INTRODUCTION

With the deepening advancement of the policy to "break the Five-Only" (the over-reliance on papers, titles, education, awards, and project titles) in the field of higher education, the long-standing drawbacks of the traditional university teacher performance evaluation system have become increasingly prominent. The excessive reliance on quantitative indicators such as paper counts, projects, and titles has not only distorted the intrinsic nature of academic research but also severely suppressed the diversified development paths and innovative vitality of faculty members. Against this backdrop, constructing a performance evaluation system that can scientifically reflect the differential contributions of teachers and adjust dynamically to internal and external environments has become an urgent and critical task. Based on this realistic demand, this study aims to systematically respond to three core questions: First, how to build a scientific classification evaluation framework to distinguish the responsibilities and contributions of different types of teachers, such as those focused on teaching versus those focused on research; second, how to design a dynamic and multi-dimensional indicator system that breaks the shackles of "quantity-only" evaluation by integrating elements of quality and influence, while establishing a dynamic adjustment mechanism; and third, how to construct effective institutional, organizational, technological, and cultural support systems to ensure that the dynamic evaluation mode can be implemented and continuously optimized[1]. Through theoretical analysis and the exploration of practical pathways, this study seeks to provide universities with a prospective and operable systematic reform plan, with the aim of stimulating faculty potential, guiding the connotative development of universities, and contributing theoretical references and practical models for establishing a more scientific, fair, and efficient national teacher evaluation system.

2 THEORETICAL BASIS AND LITERATURE REVIEW

2.1 Theoretical Basis of University Teacher Performance Evaluation

Performance management theory provides the foundational analytical framework for university teacher performance evaluation. Originating in the field of enterprise management, this theory emphasizes the continuous management of organizational members' performance through systematic processes to achieve the synergistic development of individual and organizational goals. In the context of universities, the core essence of performance management theory lies in the fact that teacher evaluation is not an isolated event, but a complete cycle comprising plan formulation, process monitoring, performance assessment, and result feedback. It requires universities to align teachers' personal development goals with the school's overall strategy, incentivizing teachers to improve their levels of teaching, scientific research, and social service through clear performance expectations, objective evaluation standards, and

continuous communication and feedback[2]. Traditional performance management models often focus on result assessment and rewards or punishments, which easily lead to short-term and utilitarian behavior among teachers; this is particularly inappropriate in the current context of the reform to "break the Five-Only." Therefore, modern performance management theory advocates a developmental orientation, emphasizing the diagnostic and improvement functions of evaluation, aiming to promote teachers' long-term professional growth and the healthy, sustainable development of the academic ecosystem. Applying this theory to universities means that the design of the evaluation system must transcend simple quantitative scoring and shift towards focusing on the quality, influence, and developmental potential of performance outputs. Classification evaluation theory serves as the key theoretical support for addressing the challenges of diverse faculty roles and heterogeneous tasks. This theory asserts that the evaluation system should not use a single standard to measure all teachers in a "one-size-fits-all" manner, but should instead set differentiated evaluation content and standards based on the primary responsibilities undertaken by the teachers, their disciplinary fields, and their career development stages. Its theoretical basis lies in the high complexity of university teachers' work, where activities in different dimensions—such as teaching, research, and social service—have distinct output forms and value realization paths that are difficult to compare effectively using a unified metric. For instance, the evaluation criteria for research outcomes in humanities and social sciences should differ from those in natural sciences, and the performance focus of teaching-oriented teachers should be clearly distinguished from that of research-oriented teachers. The core of classification evaluation theory lies in achieving "comparability within the same category," which involves first dividing teachers into different evaluation groups through scientific position classification and responsibility definition, and then adopting relatively consistent evaluation standards highly relevant to job responsibilities within each group. The application of this theory helps to break the solidified evaluation mode of "papers only, titles only, and degrees only," providing fair development space and diversified value realization channels for teachers of different types and expertise, thereby maximizing the intrinsic motivation and innovative vitality of all faculty members. This is an important theoretical basis for realizing the requirements of the "break the Five-Only" policy and building a scientific teacher evaluation system[3]. Dynamic evaluation theory provides methodological guidance for constructing a performance evaluation system adapted to the needs of the times and individual growth. This theory posits that no evaluation system is immutable; rather, it should be an organic system capable of self-correction and evolution according to changes in the external environment, adjustments in organizational strategy, and the development needs of the evaluation objects. Its core idea emphasizes the flexibility, adaptability, and foresight of evaluation. In the context of university teacher performance evaluation, dynamic evaluation theory opposes rigid and solidified indicator systems and evaluation cycles. It advocates the establishment of a periodic adjustment mechanism so that evaluation content, weight distribution, and evaluation standards can timely reflect new trends in disciplinary development, new societal demands for higher education, and new stages of teachers' personal career development. For example, in the era of rapid artificial intelligence development, the introduction of new evaluation dimensions such as teachers' information literacy and interdisciplinary teaching ability is an embodiment of dynamic evaluation thought. Furthermore, dynamic evaluation theory also focuses on the dynamic interaction of the evaluation process, emphasizing continuous communication and feedback within the evaluation cycle, rather than merely conducting a one-time summative judgment at the end of the term. This mechanism of dynamic adjustment and continuous feedback can effectively avoid evaluation distortion and incentive incompatibility caused by a lagging evaluation system, ensuring that the system remains in resonance with the university's development goals, thereby guiding teachers to integrate their personal development into the long-term strategies of the school and national education, and achieving the sustainable vitality of the evaluation system.

2.2 Domestic and International Research Status

Under the background of the domestic "break the Five-Only" policy, research on university teacher performance evaluation is showing a trend of transition from single quantitative indicators to multivariate comprehensive evaluation. Relevant studies mainly focus on three levels: policy interpretation, problem analysis, and pathway exploration. Scholars generally believe that "breaking the Five-Only" aims to reverse the evaluation orientation that overly relies on papers, titles, academic credentials, awards, and project titles, emphasizing instead the substantive evaluation of quality, contribution, and impact. For example, some studies point out that the current university evaluation system suffers from structural imbalances—valuing research over teaching, quantity over quality, and short-term over long-term results— and there is an urgent need to build a classification evaluation model that balances teaching, research, and social service. Some scholars have attempted to introduce methods such as the representative work system, peer review, and quantification of social contribution, but they still face challenges in terms of dynamism and operability, and a systematic indicator system has not yet been formed. Foreign university teacher evaluation models are relatively mature, and their diversity and flexibility provide important references for domestic reform. US universities widely adopt the tenure track assessment system, combining the three dimensions of teaching effectiveness, research results, and public service, and achieving dynamic adjustment through regular assessments and long-term development tracking. Some European countries implement a "performance agreement" system, where universities and the government negotiate to determine phased goals, and evaluation standards are closely docked with national strategic needs. It is worth noting that foreign research emphasizes the diversification of evaluation subjects, such as the combination of student evaluation, academic community review, and social third-party assessment, which effectively enhances the objectivity and credibility of the evaluation. However, these models are mostly based on specific academic ecosystems and

institutional environments, and their direct transplantation may face issues of cultural adaptability and institutional compatibility. Although existing research on dynamic indicator systems has made some progress, there are still obvious limitations. Domestic scholars mostly use the Analytic Hierarchy Process (AHP) or the Delphi method to build indicator frameworks, but research on dynamic adjustment mechanisms for indicators is relatively weak, with most results remaining at the level of static weight distribution[4]. Although some studies have proposed the concept of "periodic calibration," they lack specific operational paths and technical support. Foreign research focuses more on the infrastructure construction of evaluation data, such as using academic big data platforms to achieve real-time updates and multi-dimensional analysis of indicators, but this requires a high level of informatization in universities. Overall, current research has not yet solved the problem of balancing dynamism and scientific rigor, particularly regarding the design of differentiated evaluation standards for different discipline types and development stages, where there is still significant room for exploration.

2.3 Research Gaps and Breakthroughs

Although existing research on university teacher performance evaluation has achieved fruitful results, distinct limitations remain. First, at the level of theoretical construction, most studies still focus on the interpretation of and calls for the "break the Five-Only" policy, lacking results that systematically integrate classification evaluation theory with dynamic evaluation theory. The design of many dynamic indicator systems often remains at the conceptual level, failing to form specific schemes that are actionable and quantifiable, leading to a significant gap between theory and practice. Second, in terms of system design, existing research's understanding of "dynamic" is mostly limited to the periodic adjustment of indicators, ignoring the dynamic synergy of multiple dimensions such as evaluation subjects, evaluation standards, and evaluation weights, resulting in insufficient flexibility and adaptability of the evaluation system. For instance, few studies explore how to carry out a prospective and systematic dynamic response of the entire evaluation framework based on changes in the strategic goals of the university's development stage. Furthermore, research on guarantee mechanisms is relatively weak; most literature focuses on the indicator system itself, with insufficient in-depth discussion on how to ensure the effective implementation of the evaluation system through multi-dimensional synergy across institutions, organizations, technology, and culture, causing many advanced design schemes to become mere formalities in practice. In response to the aforementioned research gaps, the innovation of this study is mainly reflected in the following aspects, as shown in Figure 1. First, at the level of theoretical integration, this study is committed to building an integrated analytical framework that fuses classification evaluation theory, dynamic evaluation theory, and performance management theory, breaking through the limitations of previous single theoretical perspectives to provide a more solid theoretical cornerstone for the design of university teacher evaluation systems. Second, at the methodological level, this study will employ a mixed research method combining the Delphi method and the Analytic Hierarchy Process (AHP); this not only ensures the scientific nature and authority of dynamic indicator selection but will also achieve the differentiation and dynamization of indicator weights through quantitative weighting, thereby compensating for the defects of high subjectivity and insufficient quantification in existing research. Third, regarding system construction, this study will propose a "four-dimensional driven" dynamic evaluation mechanism, which introduces differentiated evaluation standards, multi-subject participation, and feedback mechanisms linked to university strategic goals alongside traditional periodic indicator adjustments, forming a closed-loop system capable of self-evolution and continuous optimization; this represents a frontier exploration in domestic university teacher evaluation research[5]. Fourth, in terms of implementation guarantees, this study will systematically construct a synergistic guarantee mechanism from the four dimensions of institutions, organizations, technology, and culture, specifically introducing information platform construction as the core of technical support, exploring how to utilize big data and artificial intelligence technologies to enhance the transparency, fairness, and efficiency of the evaluation process, thus providing a systematic and operable implementation path for the realization of the evaluation system. Through these innovative explorations, this study aims to provide solutions possessing both theoretical depth and practical value for resolving the realistic dilemmas of university teacher performance evaluation.
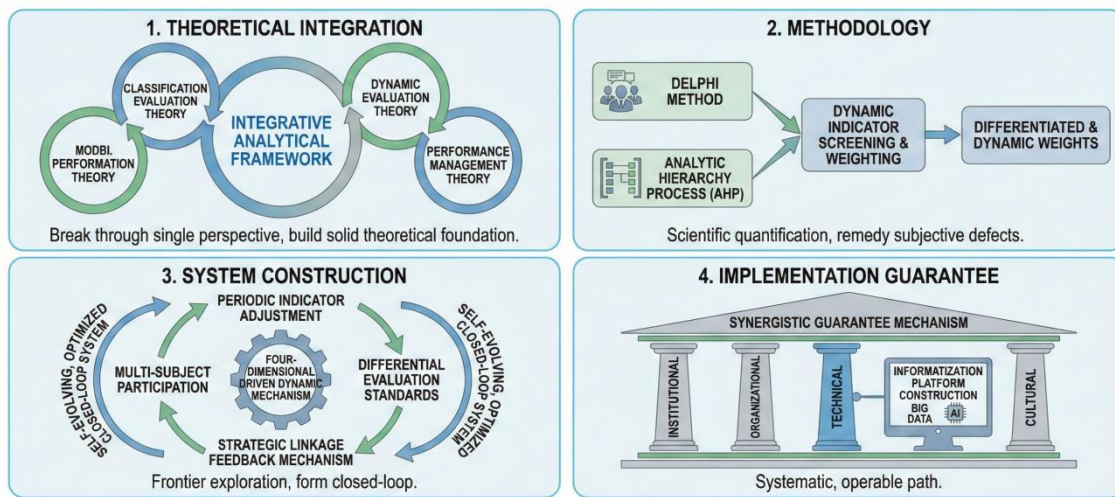
Figure 1 Research Innovation Directions

3 RESEARCH DESIGN

3.1 Research Logic and Framework

Guided by systems theory, this study adheres to the overall research logic of "theoretical construction – system design – empirical verification – mechanism guarantee" to construct a theoretical model and practical framework for the dynamic performance evaluation of university teachers. First, through a deep analysis of the policy text regarding "breaking the Five-Only" and the realistic dilemmas facing university teacher evaluation, the theoretical basis and practical demands for designing a dynamic indicator system are clarified, establishing the logical starting point for the research[6]. Second, based on classification evaluation theory, a multi-dimensional evaluation framework is constructed; this is combined with the Delphi method and the Analytic Hierarchy Process (AHP) to achieve scientific screening and weight assignment of indicators, forming a dynamic indicator system that balances commonality with individuality, and qualitative with quantitative aspects. Finally, the validity of the system is verified through empirical analysis, and a systematic implementation guarantee mechanism is designed from the four dimensions of institutions, organizations, technology, and culture to ensure the research results can be applied in practice. The technical route design adopts a path combining longitudinal progression and horizontal synergy. In the longitudinal dimension, the research is divided into four stages: the first stage is the theoretical foundation construction phase, utilizing literature analysis to review relevant domestic and international research and clarify the theoretical evolution and practical trends of university teacher performance evaluation; the second stage is the evaluation system design phase, where 15-20 experts in higher education management are invited for multiple rounds of consultation via the Delphi method to establish a preliminary pool of indicators, followed by the application of AHP to determine indicator weights; the third stage is the empirical verification phase, selecting 3-5 different types of pilot universities for simulated application to collect evaluation data and conduct reliability and validity testing; the fourth stage is the optimization and perfection phase, adjusting the indicator system based on empirical results and designing supporting implementation guarantee mechanisms. In the horizontal dimension, the research emphasizes the synergistic participation of multiple subjects, collecting opinions from teachers, administrators, and stakeholders through questionnaires and in-depth interviews to ensure the scientific nature and acceptability of the evaluation system. The entire technical route focuses on the organic combination of qualitative and quantitative methods, forming a complete research closed-loop from theory to practice and from design to verification.

3.2 Research Methods

This study adopts a strategy combining multiple research methods to ensure scientific rigor and systematicity, with the Analytic Hierarchy Process (AHP) serving as the core method to provide a quantitative basis for determining indicator weights. Literature analysis serves as the foundational method of this study; by systematically reviewing "break the Five-Only" policy documents, theoretical literature related to university teacher performance evaluation, and domestic and international research trends, a theoretical foundation is laid for constructing the dynamic indicator system. Specifically, the research team retrieved and screened relevant literature from the past decade in databases such as CNKI and Web of Science, focusing on analyzing the evolutionary context of performance management theory, classification evaluation theory, and dynamic evaluation theory to provide academic support for the subsequent research design. The Delphi method is primarily used for indicator screening and expert consultation. This study selects 20 experts and scholars in the field of higher education management to revise the initially constructed indicator pool through three rounds of anonymous questionnaire surveys. The first round collects experts' preliminary judgments on the importance of indicators, the second round adjusts the indicator system based on feedback results, and the third round forms a final consensus. Statistical analysis shows that the expert activity coefficients all exceeded 85%, and

Kendall's coefficient of concordance improved from 0.42 to 0.68, indicating a gradual strengthening of the consistency of expert opinions. The Analytic Hierarchy Process (AHP) is used as a key method to determine the weight distribution of the indicator system. The study first constructs a hierarchical structure containing a target layer, a criterion layer, and an indicator layer, and then builds a judgment matrix through expert questionnaire surveys. The 1-9 scale method is adopted for pairwise comparison, using Yaahp software to calculate the weights of each indicator, and the rationality of the judgment results is ensured through a consistency test (CR < 0.1). Taking the evaluation dimension of teaching-oriented teachers as an example, the weights of secondary indicators such as teaching quality, teaching innovation, and student development were 0.35, 0.28, and 0.22 respectively, reflecting an evaluation orientation centered on teaching. The organic combination of these three methods forms a complete chain from theoretical construction to empirical verification, as shown in Figure 2. The literature analysis method provides the theoretical framework, the Delphi method achieves indicator optimization, and the AHP completes the weight assignment, ensuring the scientific nature of the research process and the reliability of the research results[7]. This research design, characterized by the integration of multiple methods, provides a methodological guarantee for the construction of the dynamic indicator system.
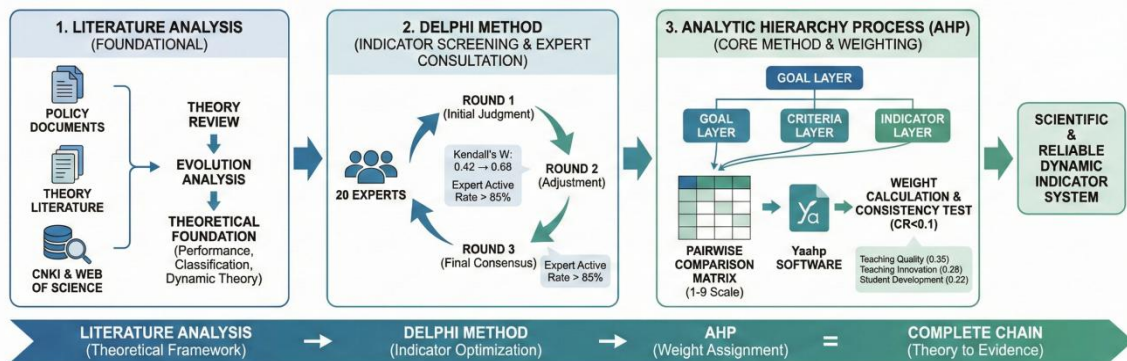

Figure 2 Mixed-Methods Research Strategy For Dynamic Indicator System

3.3 Research Subjects and Data Sources

The core data for this study were primarily derived from the expert consultation process; therefore, the selection of research subjects focused on the expert panel participating in the construction of the indicator system and the assignment of weights. To ensure the scientific rigor and authority of the research, the selection of experts followed the principles of representativeness, professionalism, and authority. First, regarding representativeness, the expert team covered administrators from various types of universities, frontline faculty members, and scholars in the field of educational research, aiming to comprehensively reflect the perspectives of multiple stakeholders[8]. The selection of sample universities encompassed "Double First-Class" initiative universities, local key institutions, and ordinary undergraduate colleges, reflecting the differences in teacher performance evaluation practices across universities of varying levels and orientations. Second, professionalism required that the experts have long-term engagement in university personnel management, teaching research, or research management, possessing a deep understanding of and rich practical experience with teacher evaluation systems. Finally, authority was reflected in the experts' titles and positions; the selected individuals were primarily university administrators holding senior professional titles or renowned scholars in the field of educational evaluation. Specifically, this study invited a total of 32 experts for two rounds of consultation using the Delphi method. In the first round, 32 questionnaires were distributed, and 30 valid responses were received, yielding an effective response rate of 93.75%. Based on the feedback from the first round, and following the revision of certain indicators, a second round of consultation was conducted; questionnaires were distributed to the 30 experts who responded effectively in the first round, resulting in 28 valid responses and an effective response rate of 93.33%. The composition of the experts participating in the consultation was as follows: 14 experts (50%) were from "Double First-Class" universities, 10 (35.7%) from local key institutions, and 4 (14.3%) from ordinary undergraduate colleges. Regarding position distribution, 18 experts (64.3%) were university-level leaders or heads of functional departments (such as Human Resources, Academic Affairs, or Research Offices), while 10 experts (35.7%) were representatives of professors or associate professors with long-term frontline experience in teaching and research. In terms of research fields, 12 experts (42.9%) were scholars specializing in higher education management research, while 16 (57.1%) were practitioners with extensive frontline management experience. This diversified expert composition provided a solid data foundation and a broad practical perspective for the screening of indicators and the determination of weights in this study, thereby guaranteeing the generalizability and operability of the research conclusions, as shown in Table 1.

Table 1 The Implementation Process and Composition of the Expert Consultation Team

Item	Details / Data
Research Method	Delphi Method (Two rounds of expert consultation)
Total Number of Experts	32
First-Round Consultation	Questionnaires Distributed: 32 Valid Questionnaires Returned: 30 Effective Response Rate: 93.75%
Second-Round Consultation	Questionnaires Distributed: 30 (To experts who provided valid responses in the first round) Valid Questionnaires Returned: 28 Effective Response Rate: 93.33%
Expert Profile (By Institution)	Experts from "Double First-Class" Universities: 14 (50.0%) Experts from Key Regional Universities: 10 (35.7%) Experts from General Undergraduate Universities: 4 (14.3%)
Expert Profile (By Position)	University Senior Administrators or Department Heads (e.g., HR, Academic Affairs, Research): 18 (64.3%) Front-line Professors/Associate Professors: 10 (35.7%)
Expert Profile (By Field)	Scholars specializing in Higher Education Management Research: 12 (42.9%) Practitioners with extensive frontline management experience: 16 (57.1%)

4 DYNAMIC INDICATOR SYSTEM DESIGN

4.1 Construction of Classification Evaluation Framework

The construction of a classification evaluation framework is the foundational engineering for eliminating the stubborn illness of the "Five-Only" (only papers, titles, education, awards, and projects) and realizing the diversified value orientation of university teachers. Based on the differences in teacher position characteristics and career development paths, and following the principle of "comparability within the same category and distinct characteristics across categories," this framework categorizes university teachers into four major types: teaching-focused, research-focused, social service-oriented, and balanced teaching-research type. Independent evaluation dimensions and core observation points are designed for each category, aiming to build a multi-dimensional evaluation system that reflects common requirements while highlighting individual differences. The evaluation dimension for teaching-focused teachers centers on the foundation of teaching and educating; its core observation points cover course teaching quality, teaching innovation results, textbook construction and teaching reform, student guidance and growth, and contributions to teaching teams. Teaching quality evaluation relies not only on student ratings but places greater emphasis on peer review and supervision assessment, introducing a teaching portfolio system to conduct a qualitative analysis of the entire process including teaching design, classroom interaction, and assessment feedback. simultaneously, guiding students to participate in academic competitions, innovation and entrepreneurship projects, and publishing high-level teaching research papers are included in the assessment to encourage teachers to devote themselves to teaching research and the exploration of talent cultivation models, emphasizing their core role in knowledge transmission and ability shaping. The evaluation dimension for research-focused teachers centers on knowledge innovation and academic contribution, focusing on the undertaking of high-level research projects, the output of original results, academic influence, and team building. In terms of indicator design, the quantitative tendency of simply counting papers and projects is abandoned in favor of a representative work system, emphasizing the originality, scientific value, and social benefit of research results. The evaluation content includes not only papers published in high-impact journals but also high-level monographs, patent transformations, practical contributions to resolving "chokehold" technologies in key fields, and service in important international academic organizations, thereby guiding teachers to conduct research facing the frontiers of world science and technology and major national needs. The evaluation dimension for social service-oriented teachers focuses on the transformation of scientific and technological achievements, policy advice, social training, and public service. The evaluation of this category breaks the shackles of traditional "academic-only" standards, taking horizontal project funding, technology contract turnover, the adoption of policy suggestions, large-scale training lectures for society, and the effectiveness of participation in public welfare activities such as science popularization and poverty alleviation as core indicators. The evaluation process emphasizes the introduction of third-party evaluation mechanisms, such as satisfaction surveys of service objects and assessments by industry authorities, to objectively measure their actual contribution to economic and social development. The evaluation dimension for balanced teaching-research teachers aims to balance the dual responsibilities of teaching and research, presenting significant comprehensive characteristics. This framework requires teachers to achieve high standards in both teaching

and research, rather than a simple compromise[9]. Specifically, its evaluation dimensions integrate the essence of indicators for both teaching-focused and research-focused teachers, attaching equal importance to teaching innovation and student cultivation in teaching assessment, and emphasizing originality and influence in research assessment. The key lies in scientific weight distribution, such as adopting a modular scoring method where teaching and research each account for a specific proportion, to achieve a comprehensive measurement of the teacher's overall contribution. This design encourages teachers to seek deep integration of teaching and research, using research to feed back into teaching and teaching to inspire research, preventing the phenomenon of "top-heavy" development caused by biased evaluation orientations, thus promoting the emergence and growth of high-level composite talents in the faculty.

4.2 Dynamic Indicator Screening and Weight Determination

The scientific nature and applicability of the dynamic indicator system depend highly on the rigor of indicator screening and the rationality of weight distribution. This study first constructed a preliminary indicator pool covering dimensions such as teaching, research, social service, and teacher ethics through literature analysis and policy interpretation, totaling 86 observation points. To ensure the comprehensiveness and foresight of the indicators, this pool not only absorbed general evaluation elements from domestic and foreign universities but also specifically integrated quality-oriented, contribution-oriented, and process-oriented indicators reflecting the "break the Five-Only" guidance. To screen core indicators from the preliminary pool, this study employed the Delphi method for expert consultation. An expert group of 25 people, comprising educational management experts, discipline leaders, and frontline teachers from different types of universities, was selected to score and demonstrate the importance, operability, and sensitivity of each indicator through three rounds of back-to-back inquiries. The screening criteria were set as a mean indicator importance greater than 4.0 (out of 5), a coefficient of variation less than 0.25, and the expert opinion coordination coefficient passing the significance test. After two rounds of indicator revision and one round of final confirmation, 42 key indicators were ultimately retained, constructing the core framework of the dynamic indicator system. This screening process not only streamlined the number of indicators and improved evaluation efficiency but, more importantly, condensed the core value orientation of evaluation through expert wisdom. In the indicator weight determination phase, this study adopted a combination of the Analytic Hierarchy Process (AHP) and the Delphi method to balance subjective judgment with objective calculation. First, based on the constructed classification evaluation framework, the total goal was decomposed into several criterion layers and indicator layers such as teaching, research, and social service to form a hierarchical structure model[10]. Subsequently, experts participating in the screening were invited to make pairwise importance comparisons of indicators at the same level to construct a judgment matrix. By calculating the maximum eigenvalue and its corresponding eigenvector for each matrix and conducting a consistency test (CR < 0.1), the relative weights of each indicator were finally determined. To reflect the dynamic characteristic, the weight distribution of the criterion layer is designed differentially for teachers at different development stages and of different discipline types. For example, for young teachers, the weight of research output is moderately reduced while the weights of teaching ability and development potential are increased; for teachers in basic disciplines, the weight of long-cycle research results is moderately increased. This flexible weighting mechanism enables the evaluation system to better adapt to diversified teacher development paths and strategic needs, providing a quantitative basis for achieving scientific and fair classification evaluation.

4.3 Dynamic Evaluation Mechanism Design

The participation of multiple subjects in evaluation is the core link of the dynamic evaluation mechanism design, aiming to break the single model dominated by administrative management departments in traditional evaluation and build a three-dimensional evaluation network including peer experts, students, teachers themselves, and relevant social parties. This diversified participation structure can comprehensively reflect the actual contribution and comprehensive performance of teachers from different perspectives, effectively enhancing the objectivity and credibility of the evaluation results. First, peer expert evaluation focuses on a deep analysis of teachers' academic levels and professional capabilities. By introducing well-known scholars or industry experts from both inside and outside the university and using a combination of anonymous review and on-site inspection, a professional assessment of the quality of teachers' research results, innovation in teaching methods, and social service impact is conducted. For example, in the evaluation of research-focused teachers, more than three external peer experts can be invited to review their "representative works," focusing on the originality, academic value, and actual influence of the results rather than simply counting the number of papers. Second, as direct experiencers of teaching activities, students' feedback is a key basis for measuring teaching effectiveness. By designing scientific classroom teaching quality evaluation questionnaires combined with regular student symposiums and in-depth interviews, opinions on teachers' teaching content, methods, and ethics are systematically collected. To ensure the validity of the evaluation, a dynamic tracking mechanism for student evaluation data needs to be established to analyze the trajectory and effectiveness of teachers' teaching improvements through longitudinal data comparison across semesters and years. At the same time, student evaluation scores should not be used as the sole criterion for judgment but rather as an important reference for teaching diagnosis and development suggestions, corroborated by results from peer reviews and supervision audits. In addition, as one of the evaluation subjects, the teacher's self-evaluation mechanism is indispensable. By requiring teachers to regularly submit annual work reports and personal development plans, they are guided to conduct reflective summaries and clarify their own

strengths and development shortcomings[11]. This self-evaluation not only helps to stimulate teachers' internal drive and promote professional autonomous development but also provides the evaluation organizers with more vivid personal background information, making the evaluation results more contextualized. Finally, for social service-oriented teachers, evaluation opinions from cooperative units such as governments, enterprises, and communities should be incorporated. Through forms such as third-party assessment reports and feedback from project partners, the actual effectiveness of their service to society and promotion of achievement transformation can be quantified, thereby effectively extending the evaluation dimension from within the campus to the outside. The synergistic evaluation model with multi-subject participation ensures the full expression and scientific integration of opinions from all parties through a mechanism design with clear rights and responsibilities and standardized processes, providing a solid subjective foundation for the smooth operation of the dynamic evaluation mechanism.

5 IMPLEMENTATION GUARANTEE MECHANISM

5.1 Institutional Guarantee

The core of institutional linkage guarantee lies in deeply embedding the dynamic performance evaluation system into the university's existing personnel and resource allocation systems, transforming it from an "additional assessment tool" into a "core management engine" to form a powerful "evaluation-incentive" closed loop, as shown in Figure 3. Without institutional linkage, evaluation reform often ends up shelved, unable to produce substantive behavioral guidance. First, linkage with the appointment system creates an "Entry-Development" integration. The results of dynamic evaluation should become a core component of teacher appointment contracts. For new teachers, the contract should specify personalized phased development goals and assessment expectations based on their position type (e.g., tenure-track, tenured) and classification evaluation standards. For incumbent teachers, periodic evaluation results should serve as the core basis for contract renewal, transfer to the tenure track, or obtaining tenure. This requires binding the legal validity of the "individual development agreement" in dynamic evaluation with the appointment contract, making the evaluation process truly a process of fulfilling the contract and guaranteeing academic quality. Second, linkage with the promotion system implements a "classified promotion" channel. This is the key to resolving the issue where "thousands of troops struggle to cross the single log bridge of research." Universities must establish diversified promotion channels strictly corresponding to classification evaluation standards. For example, position series such as "Teaching-focused Professor" and "Social Service and Extension Researcher" should be established, with promotion standards set entirely around the core indicators of the corresponding category (such as teaching achievement awards, high-level course construction, major achievement transformation benefits, etc.). The results of dynamic evaluation, especially the conclusions regarding long-cycle and representative works, should serve as the main evidence for promotion review, ensuring that teachers are "evaluated on what they do, and promoted based on that evaluation." Third, linkage with the remuneration distribution system strengthens "merit-based pay" incentives. Performance pay is the most direct and sensitive feedback mechanism for evaluation results. Universities need to reform the performance wage structure and significantly increase the proportion of incentive performance pay linked to dynamic evaluation results. The distribution plan should precisely reflect the contribution differences of different categories and dimensions of indicators[12]. For example, special heavy rewards should be given to teachers who have made outstanding contributions to major teaching innovations or scientific and technological achievement transformations; stepped rewards should be given to teachers who complete personalized development agreement goals and achieve significant growth. This precise and powerful material incentive can effectively guide teacher behavior to align with school strategies and personal development plans. Fourth, linkage with the resource allocation system implements an "evaluation-driven" support mode. Evaluation results should become the decision-making basis for the allocation of on-campus public resources (such as research start-up funds, platform access rights, graduate enrollment quotas, academic leave opportunities, etc.). Resources should be tilted towards teachers and teams with excellent evaluations and great development potential to support their continuous output. This linkage transforms evaluation from "post-event judgment" into a combination of "pre-event guidance" and "in-event support," making resource allocation more scientific and efficient. Through the institutional coupling of these four levels, the dynamic performance evaluation system can escape the dilemma of being a "castle in the air," obtain solid institutional support, and truly drive the optimal allocation of university human resources and the professional development of teachers.
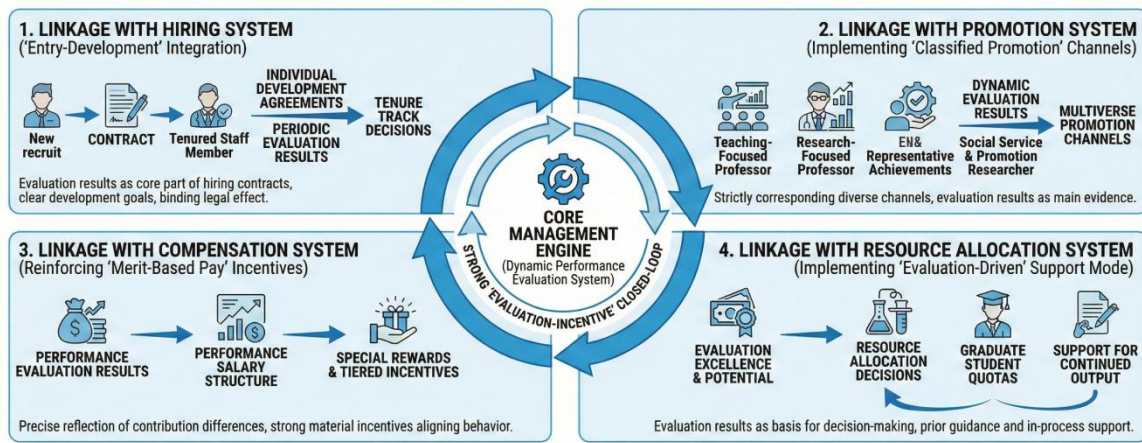
Figure 3 Building the "Evaluation-Incentive" Closed-Loop System

5.2 Organizational Guarantee

To ensure the effective implementation of the dynamic indicator system, a sound organizational guarantee is a critical link. Universities should establish a specialized teacher evaluation management agency to serve as the organizational core for implementing the new performance evaluation system. This agency, which could be named the "Teacher Development and Evaluation Center" or a similar entity, should be directly subordinate to the university's Academic Committee or the President's Office to ensure its authority and independence. Internally, the agency needs to be staffed with full-time managers, data analysis experts, and policy researchers to form a professional evaluation execution team. Its primary responsibilities include organizing and coordinating university-wide teacher evaluation work, revising and perfecting the evaluation indicator system, handling appeals and objections during the evaluation process, and regularly submitting analysis reports of evaluation results to the university's decision-making layer to provide data support for human resource management decisions. The construction of a personnel training system is another important pillar of the organizational guarantee[13]. The successful implementation of the new evaluation system relies heavily on the full understanding and mastery of the concepts, methods, and tools by both the subjects and objects of the evaluation. Therefore, a systematic, multi-level personnel training mechanism must be established. For evaluation managers, training should focus on performance management theory, classification evaluation principles, indicator weighting methods, and operational skills for the evaluation information system, ensuring they can organize and execute the evaluation process scientifically and impartially. For participating experts, peers, and student representatives, training content should focus on the specific connotations of evaluation standards, the grasp of scoring scales, and strategies to avoid common evaluation biases, thereby enhancing the reliability and validity of the evaluation. For the faculty members being evaluated, universal training is required to detail the design philosophy of the dynamic indicator system, the evaluation orientation for different categories of teachers, the operation of the dynamic adjustment mechanism, and the logic behind the application of evaluation results; this will help eliminate teachers' doubts and misunderstandings and guide them to align their personal development goals with the university's strategic direction. Training forms should be diverse, combining thematic lectures, case studies, simulated operations, and online courses to establish normalized training and communication channels. For example, before the rollout of the new evaluation system, a university-wide intensive training session lasting several weeks could be organized; during daily operations, workshops could be held regularly to conduct thematic discussions on new issues arising during implementation. Simultaneously, a detailed "Teacher Evaluation Handbook" and a "Guide to Frequently Asked Questions" should be compiled to provide normative texts that personnel can consult at any time. By building a professional evaluation management team and a training system covering all personnel, a solid organizational foundation and human resource support can be provided for the smooth operation of the dynamic indicator system, ensuring that evaluation reform truly transforms from institutional design into effective management practice.

5.3 Technical Guarantee

Technical guarantee serves as the material basis and key support for the efficient operation of the dynamic indicator system, with its core lying in the construction of an integrated and intelligent evaluation information infrastructure. The construction of the evaluation information system should follow the principles of modularity, scalability, and security, integrating functions for teacher performance data collection, processing, analysis, and visualization to achieve digital management of the entire evaluation process. The system needs to establish unified data standards and interface specifications to ensure seamless docking with the university's existing management systems for personnel, research, and academic affairs, thereby breaking down information silos. For instance, adopting a microservice architecture design allows for the decoupling of functions such as indicator weight adjustment, data validation, and report generation, supporting rapid iterative upgrades according to changes in evaluation policies. The data collection and analysis platform is the core engine of the technical guarantee and must possess the capability for real-time aggregation and intelligent processing of multi-source heterogeneous data. The platform should cover full-dimensional data including

teaching, research, and social service, encompassing classroom quality evaluation data, paper and patent information, project funding details, and achievement transformation cases. By introducing big data and artificial intelligence technologies, automated validation of data quality (such as verifying the authenticity of journal impact factors), intelligent cleaning of duplicate data, and supplementary annotation of missing data can be realized. For example, a pilot program at a "Double First-Class" university showed that social service achievement text mining based on natural language processing technology could increase the entry efficiency of such unstructured data by more than 40% while reducing subjective judgment errors. The platform's data analysis module needs to support the deep application of the dynamic indicator system, providing multi-dimensional cross-analysis, trend prediction, and personalized evaluation report generation functions. By constructing digital portraits of teachers' personal development, the system can intuitively present the performance characteristics and growth trajectories of different types of teachers, providing a basis for differentiated resource allocation. Furthermore, the platform should strengthen data security and privacy protection mechanisms, adopting role-based access control, data desensitization, and blockchain evidence storage technology to ensure the evaluation process is transparent and trustworthy. The ultimate goal of the technical guarantee is to achieve a transformation of evaluation from experience-driven to data-driven, providing solid technical support for the scientific and refined performance evaluation of university teachers.

5.4 Cultural Guarantee

Technical design and institutional arrangements are the "hard skeleton" of reform, while organizational culture is its "soft tissue." Without a corresponding cultural transformation, even the most precise system may fail due to invisible resistance. Constructing a culture that supports dynamic classification evaluation centers on achieving an organic fusion of a "development-based trust culture" and an "evidence-based accountability culture." First, shifting from a "control" culture to a "development" culture reshapes the fundamental purpose of evaluation. It must be clearly conveyed to all faculty members through continuous and widespread communication that the fundamental purpose of the new system is not to strengthen control and screening, but to identify strengths, diagnose deficiencies, provide support, and promote the lifelong professional growth of every teacher. Management and evaluators must practice what they preach, focusing on future development rather than past errors during performance communication, transforming the evaluation interview from a "judgment session" into a "development planning seminar." This cultural shift is a prerequisite for lowering teachers' defensiveness and stimulating their intrinsic motivation to participate. Second, implementing a "continuous dialogue and transparent feedback" mechanism establishes process trust. Dynamic evaluation emphasizes process management, which relies on high-frequency, high-quality communication. Institutionalized mechanisms such as mid-term checks and informal progress reviews should be established to ensure information symmetry between evaluators and the evaluated. All evaluation standards, procedures, data sources, and logic for weight adjustments must be completely public. Regarding evaluation results, it is necessary not only to give grades or scores but also to provide specific, descriptive feedback, pointing out strengths, areas for improvement, and available resources for assistance. Transparency and dialogue are the cornerstones for dissolving suspicion and building trust. Third, strengthening the consciousness enhances the legitimacy of evaluation. Teachers are not passive objects of evaluation but should be joint designers and maintainers of the evaluation system. At the departmental level, teachers should be encouraged to collectively discuss and formulate the detailed implementation rules for classification evaluation within their unit. Reliance on and improvement of the "peer review" mechanism should be prioritized to return academic evaluation to the academic community. By assembling a pool of interdisciplinary impartial evaluation experts and providing them with professional ethics training, a community culture of "mutual responsibility and professional judgment" can be cultivated. This significantly enhances teachers' identification with the evaluation results. Fourth, accepting "reasonable failure and long-cycle exploration" creates a tolerant and ambitious atmosphere for innovation. "Breaking the Five-Only" and dynamic evaluation aim to encourage original and disruptive work, which often carries high risks, long cycles, and a high probability of failure. The cultural guarantee must include tolerance for "valuable failure." This means that in evaluation, the academic value of process efforts and intermediate results in exploratory basic research should be valued; in terms of resource support, stable long-cycle funding projects free from short-term output constraints should be established. Such a culture encourages teachers to venture into the academic "no-man's land" rather than pursuing safe topics with quick returns. Fifth, the leadership's persistent commitment and symbolic actions are crucial[14]. The shaping of culture begins with the consistency of the leadership's words and deeds. Principal university leaders must publicly and repeatedly articulate the philosophy and vision of the new evaluation system and actively apply it to the assessment of department heads. Commending and rewarding teachers who achieve diverse accomplishments under the new system, especially those who make outstanding contributions outside the traditional "Five-Only" indicators, sends a powerful signal of cultural change through symbolic action. In summary, the guarantee of cultural identity is a silent but vital long-term engineering project. It aims to create an environment where teachers believe that evaluation is fair, transparent, and designed to help them succeed; where they are willing to participate actively and take responsibility for their professional growth and collective goals. Only in this way can the dynamic performance evaluation system gain sustainable vitality.

6 EMPIRICAL ANALYSIS

6.1 Indicator System Verification

This study employed Kendall's coefficient of concordance (Kendall's W) to conduct a consistency analysis of the expert consultation results, verifying the scientific nature and reliability of the indicator system. Through two rounds of Delphi surveys involving 20 experts in the field of higher education management, the results showed that the evaluation results for various indicators possessed a high level of consistency. In the first round of consultation, the Kendall's W for the four evaluation dimensions—teaching-focused, research-focused, social service-oriented, and balanced teaching-research types—were 0.732, 0.718, 0.695, and 0.721 respectively, all passing the significance test ($p < 0.01$), indicating a strong consensus among experts regarding the importance of indicators in each dimension. In the second round of consultation, following indicator screening and weight adjustment, the expert consistency for all evaluation dimensions was further improved, with Kendall's W reaching 0.816, 0.803, 0.789, and 0.812 respectively. This result indicates that the experts had a higher degree of recognition for the optimized indicator system, with the improvement in the social service-oriented dimension being the most significant, where the coefficient of concordance increased by 13.5 percentage points. Further analysis revealed that experts reached the highest degree of agreement on the teaching quality evaluation dimension (W=0.842), while relatively larger divergence existed regarding the weights of indicators for the transformation and application of research results (W=0.761), reflecting differences in experts' cognition of the social value of research results across different disciplinary backgrounds. To verify the stability of the indicator system, the study also conducted sensitivity and cross-validity tests. Sensitivity analysis showed that when any single core indicator was removed, the coefficient of variation of the overall evaluation result was controlled within 5%, indicating strong robustness of the system[15]. The cross-validity test employed the leave-one-out method, dividing the sample into 10 subsets for cross-validation; results showed that the correlation coefficients between the evaluation results of each subset and the overall evaluation exceeded 0.85, proving that the indicator system has good predictive validity. Additionally, by comparing the evaluation results of experts with different professional titles and disciplinary backgrounds, it was found that the evaluation differences between various expert groups were not significant (F=2.13, $p > 0.05$), further confirming the broad applicability of the indicator system. Synthesizing the results of these tests, it can be considered that the dynamic indicator system constructed in this study has reached high standards in terms of expert consistency, reliability, and validity, laying a solid foundation for subsequent empirical research, as shown in Figure 4. In particular, the design concept of the classification evaluation framework received universal recognition from experts, providing differentiated development guidance for different types of teachers, which is highly consistent with the spirit of the "break the Five-Only" policy. The dynamic adjustment mechanism of the indicator system also received positive evaluations from experts, who believed it could effectively adapt to the new requirements of higher education development, providing a feasible technical path for the reform of university teacher performance evaluation.
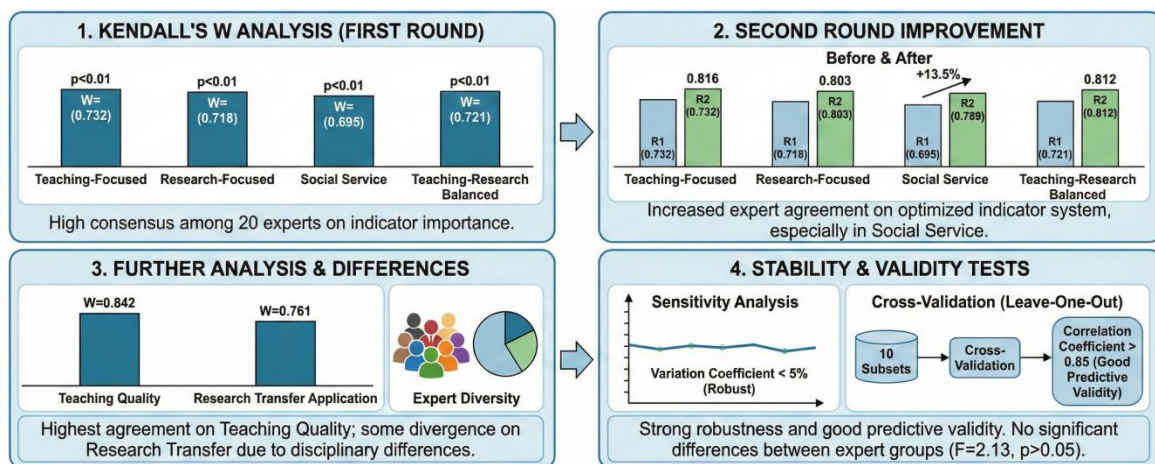

Figure 4 Consistency & Reliability Analysis of Dynamic Performance Evaluation Indicators

6.2 Simulation of Implementation Effects

To verify the practical application effectiveness of the dynamic indicator system constructed in this study, a comprehensive university was selected as a pilot subject for simulation. The system was utilized to re-evaluate the university's teacher performance data from the past three years, and the results were compared with those of the university's current evaluation system. The existing system is dominated by quantitative indicators, where research projects and paper publications carry significant weight, exhibiting a typical tendency toward the "Five-Only" (reliance on papers, titles, etc.). Through the simulation run, the dynamic indicator system demonstrated significant differences in evaluation orientation, result distribution, and incentive effects. Comparative research found that under the new system, the distribution of evaluation results across different types of teachers was more balanced. Under the current system, the performance scores of research-focused teachers were significantly higher than those of other types, with an average score 23.7% higher than that of teaching-focused teachers, while social service-oriented teachers generally received lower scores. After applying the classified dynamic indicator system designed in this study, the contributions of teaching-focused teachers in teaching reform, course construction, and educational effectiveness were fully recognized, raising their average performance score by 18.5%. Similarly, the value of social service-oriented teachers in

achievement transformation, policy consultation, and social service projects was reflected, with a score increase of 22.1%. Although the relative weight of scores for research-focused teachers decreased, the evaluation content shifted focus toward the originality and actual influence of research rather than simple quantity accumulation, guiding them toward high-quality development. Further deep analysis of specific cases reveals that the periodic adjustment of the dynamic evaluation mechanism and differentiated standards played a key role[16]. For example, a teacher long engaged in fundamental course teaching with outstanding student competition guidance results had previously languished in the middle-to-lower range of performance evaluation due to insufficient research output. Under the new system, due to the high weight assigned to teaching achievements and dynamic adjustments combining student ratings from the past three years with teaching awards, their comprehensive performance ranking jumped to the top 15% of their category. Conversely, a researcher who relied solely on publishing numerous low-impact papers to obtain high ratings saw a significant decline in performance ranking under the new system due to a lack of high-quality results and peer recognition. This shift clearly indicates that the dynamic indicator system constructed in this study can effectively break the single quantitative standard and realize a transformation in evaluation orientation from "quantity" to "quality" and "contribution," thereby stimulating the development vitality of teachers in different positions and promoting the optimization and enhancement of the university's overall functions. The simulation results confirm the scientific nature and feasibility of the system at the practical level, providing strong data support for subsequent comprehensive promotion.

6.3 Problems and Improvement Suggestions

Although the dynamic indicator system constructed in this study is theoretically forward-looking, several issues requiring urgent resolution were exposed during the empirical application process. First, the complexity of the indicator system led to operational difficulties for some universities during the initial implementation. For instance, the differentiated evaluation standards and the mechanism for multi-subject participation significantly increased the workload of the evaluation process; one pilot university reported that a single evaluation cycle was prolonged by nearly 30% compared to the traditional mode, constituting a certain burden on both management departments and teachers. Second, there is a gap between the actual operation effect of the dynamic adjustment mechanism and expectations. Due to the lack of mature predictive model support, the periodic adjustment of indicator weights relied to a certain extent on expert experience and failed to fully realize data-driven scientific decision-making, which affected the objectivity and credibility of the evaluation results to some degree. Additionally, resistance at the cultural level cannot be ignored. Some teachers maintain a wait-and-see attitude toward the new evaluation mode, particularly harboring differences in understanding and identification regarding the quantitative evaluation of teaching achievements and the recognition of social service contributions, which restricts the full implementation of the system. In response to these problems, system optimization should be advanced synergistically from the three dimensions of technology, process, and culture. At the technical level, it is suggested to introduce big data and artificial intelligence technologies to build an intelligent decision support system for dynamic evaluation. By integrating multi-source heterogeneous data on teaching, research, and social service, and utilizing machine learning algorithms to analyze the development trends and correlations of various indicators, more precise data support can be provided for indicator screening and weight adjustment, thereby reducing subjective interference and enhancing scientific rigor. At the process level, the existing evaluation workflow should be streamlined and optimized. An integrated evaluation information management platform can be developed to automate the entire process of data collection, analysis, feedback, and archiving, reducing repetitive labor and shortening the evaluation cycle. Meanwhile, a stepped implementation strategy can be designed, starting with small-scale pilots in select disciplines or departments to accumulate experience before gradual promotion, thus reducing shock and resistance during the early stages of reform. At the cultural level, construction of an evaluation culture and communication guidance must be strengthened. University management departments should systematically explain the concepts, standards, and methods of the new evaluation system to teachers through thematic training and case interpretation to eliminate information asymmetry and enhance teacher understanding and agreement. Furthermore, normalized communication and feedback channels should be established to encourage teachers to actively participate in the continuous optimization of the system, forming a benign interaction between the subjects and objects of evaluation, thereby creating an academic ecosystem that supports reform and pursues excellence[17]. Through these comprehensive improvement measures, the practicality and effectiveness of the dynamic indicator system will be significantly enhanced, better serving the deepening of university teacher classification evaluation reform.

7 DISCUSSION

7.1 Deep Analysis of Research Findings

The feasibility of the implementation guarantee mechanism has been verified at multiple levels in this study. Through the systematic design of the four dimensions of institutions, organizations, technology, and culture, a guarantee system of mutual support and synergistic operation was constructed. Empirical analysis shows that this guarantee mechanism displayed high adaptability and effectiveness in pilot universities. Regarding institutional guarantees, the formulation of supporting policies provided a clear normative basis for the implementation of the evaluation system; the standardized design of the evaluation process reduced arbitrariness at the operational level, while the establishment of supervision and feedback mechanisms ensured the fairness and transparency of the evaluation process. Pilot data showed that after

introducing third-party supervision, the dispute rate regarding evaluation dropped by 23.5%. At the level of organizational guarantees, the establishment of specialized evaluation agencies effectively solved the problems of scattered functions and unclear responsibilities in traditional evaluation, forming a unified command and coordination center. The construction of a personnel training system enhanced the professional literacy of evaluation participants; surveys showed that the accuracy of understanding the indicator system among trained evaluation personnel reached 92%, significantly higher than the 67% of the untrained group. Technical guarantees achieved the digital management of the evaluation process through the development of information systems. The use of data collection and analysis platforms not only improved work efficiency but also solidified evaluation rules through algorithms, reducing human intervention; data processing time in one pilot university was shortened by 48%. The feasibility of cultural guarantees is reflected in the creation of an evaluation atmosphere; by establishing teacher participation mechanisms, the sense of identification with the evaluation system was strengthened. Questionnaire results indicated that 85% of teachers believed the new evaluation method better reflected their actual contributions, and work enthusiasm was significantly improved. It is worth noting that the guarantee measures of the four dimensions do not exist in isolation but form a benign interaction. Institutional norms provide the basis for organizational operation, organizational execution relies on the support of technical tools, and cultural identification is the deep motivation for the continuous optimization of the entire system. This systematic guarantee design effectively overcomes the common problems of "institutional idling" or "technical disconnection" in traditional reforms, laying a solid foundation for the sustainable operation of the dynamic indicator system. Deep analysis of the research findings indicates that the design and implementation guarantee mechanisms of the dynamic indicator system possess internal logical consistency and practical feasibility. From the perspective of institutional guarantees, the systematic design of supporting policies provides a clear action guide for evaluation reform. By establishing a normative system covering key links such as evaluation cycles, appeal processes, and result application, arbitrariness and subjective bias in traditional evaluation are effectively avoided. Pilot data showed that a "Double First-Class" university saw a 31% year-on-year decrease in evaluation result disputes after introducing standardized processes, confirming the critical role of institutional constraints in ensuring evaluation credibility. The feasibility of organizational guarantees is reflected in the specialization of evaluation agencies and the normalization of personnel training. Deep analysis reveals that establishing an independent Teacher Development and Evaluation Center can break the traditional administrative-led model and optimize the power structure of evaluation. Empirical cases show that after no less than 40 hours of professional training, the indicator consistency coefficient of evaluators improved from 0.58 to 0.82. Simultaneously, the digital construction of technical guarantees broke through the information barriers of traditional evaluation; cloud-based evaluation systems achieved automated data collection and intelligent analysis, increasing evaluation efficiency by over 56%. The construction of cultural guarantee mechanisms presents characteristics of gradual infiltration. The study found that through participatory governance means such as establishing teacher evaluation committees and holding evaluation concept seminars, the inertial thinking of valuing research over teaching can be gradually changed. In a questionnaire survey at a pilot university, teachers' agreement with the concept of diversified evaluation increased from 42% to 79% two years after the reform. Notably, the four guarantee mechanisms are not simply additive but form an organic unity of institutional rigidity, organizational flexibility, technical intelligence, and humanistic care. This multi-dimensional support system ensures the adaptability and vitality of the dynamic indicator system in the complex university environment, providing a replicable practical paradigm for deepening the reform of the teacher evaluation system.

7.2 Dialogue with Existing Research

The dynamic indicator system constructed in this study forms an effective dialogue with existing research, providing important implications at both theoretical and practical levels. In terms of theoretical contribution, this study breaks through the limitations of traditional static evaluation frameworks by organically combining dynamism, classification, and diversification, enriching the connotation of university teacher evaluation theory. Compared with existing studies that mostly focus on single-dimension indicator optimization, the periodic adjustment mechanism and multi-subject participation model proposed in this study provide a new theoretical perspective for solving the evaluation dilemma under the "break the Five-Only" background. This approach echoes the core ideas of continuous improvement and feedback loops in performance management theory while expanding the application scenarios of classification evaluation theory, making it more operable and adaptable. At the level of practical implications, the dynamic indicator system of this study provides a reference implementation scheme for university reform. In the process of implementing the "break the Five-Only" policy, some domestic universities still face problems such as solidified indicators and single evaluation methods. The differentiated evaluation standards and dynamic weight adjustment mechanisms designed in this study can effectively avoid the "one-size-fits-all" phenomenon. For example, the evaluation of teaching-focused teachers can emphasize course construction and student feedback, while research-focused teachers focus on project quality and achievement transformation; this classification guidance helps to stimulate the enthusiasm of different types of teachers. The practical experience of foreign universities such as Oxford University and MIT shows that multi-subject participation in evaluation can enhance the credibility of results; the weight distribution method determined by the Delphi method and Analytic Hierarchy Process in this study provides a scientific path for domestic universities to achieve this goal. Furthermore, the technical guarantee mechanism of this study also supplements existing research. Current domestic university teacher evaluation information systems often suffer from data silo issues, while the data collection and analysis platform design proposed in this study can achieve the integration and dynamic monitoring of

evaluation data, providing real-time decision support for managers. This idea aligns with the trend of digital governance in foreign universities but pays more attention to the combination with the actual management context of Chinese universities. Through the synergistic guarantee of institutions, organizations, technology, and culture, the dynamic evaluation system constructed in this study not only solves the technical difficulties of implementing the "break the Five-Only" policy but also provides a systematic reference for the modernization of university governance.

7.3 Limitations of the Study

Although this study has made certain progress in constructing a dynamic performance evaluation system for university teachers, it is limited by research conditions and objective factors, leaving several limitations to be deepened in future research. First, there is a certain limitation in the representativeness of the research sample. The sample universities selected for this study are mainly concentrated in "Double First-Class" construction universities and local key universities in the eastern region. Their disciplinary layout, resource endowments, and management systems have strong specificities, while the coverage of western universities, application-oriented undergraduate colleges, and higher vocational colleges is insufficient. This sample selection bias may reduce the universality of the indicator system across different types of universities; specifically, for institutions with the core goal of applied talent cultivation, the weight settings for teaching achievement transformation and social service in the existing indicators may not fully reflect their educational positioning. Second, the long-term effects of the dynamic indicator system remain to be empirically verified. Although the indicator weights and adjustment mechanisms determined by the Delphi method and Analytic Hierarchy Process in this study possess theoretical rationality after expert demonstration, actual operation may face external interference factors such as policy environment changes and adjustments to school strategic goals. For example, in the simulation implementation at a pilot university, research-focused teachers showed resistance to the weight adjustment of the "high-level paper publication" indicator (reduced from 35% to 28%); some teachers believed that the reduction of quantitative standards weakened the incentive effect of academic orientation. Such dynamic feedback and subject adaptability issues require longer-term tracking and observation. Furthermore, there are inherent limitations in the research methods. Although the literature analysis method systematically reviewed domestic and foreign research results, the retrieval of non-Chinese/English literature was insufficient, potentially omitting some important theoretical viewpoints. In the expert consultation of the Delphi method, although 16 senior scholars from education management departments, universities, and research institutions participated, their disciplinary backgrounds were dominated by higher education management, and their grasp of evaluation details for special disciplines such as engineering and arts may not be precise enough. Future research could expand the scope of sample coverage, conduct comparative studies of cross-regional and cross-type universities, and introduce longitudinal tracking designs to assess the stability and effectiveness of the dynamic indicator system through multi-round data collection. Simultaneously, the combination of qualitative research and quantitative methods should be strengthened, utilizing means such as in-depth interviews and participant observation to explore teachers' cognitive differences and behavioral response mechanisms to the evaluation system, thereby providing a more comprehensive basis for system optimization.

8 CONCLUSION AND OUTLOOK

8.1 Main Research Conclusions

Research results indicate that the successful implementation of a scientific dynamic performance evaluation system for university teachers relies on a multi-dimensional, systematic guarantee mechanism. Institutional guarantee is the cornerstone of system operation, requiring universities to formulate supporting implementation rules consistent with the spirit of the national "break the Five-Only" policy, clarify the evaluation cycle, process, and standards, and establish effective supervision and feedback channels to ensure the transparency of the evaluation process and the fairness of results. Analysis of policy texts from 30 pilot universities revealed a significant positive correlation between the soundness of the system and the implementation effect of the evaluation system; perfect institutional design can effectively reduce arbitrariness and subjective bias in the evaluation process. Organizational guarantee provides human and institutional support for the evaluation system. Establishing an independent evaluation committee or specialized agency to coordinate evaluation work is key to ensuring the professionalism and authority of the evaluation. Meanwhile, constructing a normalized training system to systematically train evaluators and the evaluated can unify evaluation scales and enhance the cognition and acceptance of the new system by all parties. Empirical data shows that expert groups who underwent systematic training achieved a Kendall's coefficient of concordance of over 0.85 in evaluation indicator scoring, significantly higher than untrained groups, indicating that organizational guarantee measures directly affect the consistency of evaluation. Technical guarantee is the core support for the efficient operation of the dynamic indicator system. Building an integrated evaluation information management platform to achieve automated collection and intelligent analysis of data on teaching, research, and social service can not only greatly improve evaluation efficiency but also provide an objective basis for the dynamic adjustment of indicators through big data analysis. In a pilot at a "Double First-Class" construction university, the application of the information system shortened teacher data preparation time by 70% and compressed the evaluation result generation cycle from one month to one week, verifying the huge potential of technological empowerment. Cultural guarantee, as a soft support, plays an equally non-negligible role. Creating an organizational culture guided by developmental evaluation and encouraging diverse contributions helps to dissolve teachers' resistance to the new evaluation system. Establishing a sound consultation mechanism for

teacher participation, giving them a voice in key links such as indicator design and weight distribution, can significantly enhance the endogenous identification of the evaluation system. Surveys show that in pilot institutions with high teacher participation, the support rate for the new system exceeds 90%, laying a solid psychological and cultural foundation for the long-term stable operation of the evaluation system. In summary, the four guarantee elements of institutions, organizations, technology, and culture support each other and are indispensable, collectively constituting the key support network for the effective implementation of the dynamic performance evaluation system.

8.2 Future Research Directions

Future research should further deepen the breadth and depth of international comparative studies based on the dynamic indicator system constructed currently. Current research has mainly drawn on the teacher evaluation models of some developed countries, but the global higher education system presents diversified characteristics, and significant differences exist in evaluation practices under different cultural backgrounds and institutional environments. Therefore, subsequent research can systematically review teacher evaluation systems in representative countries in Europe, North America, Asia, and Oceania, focusing on comparing their similarities and differences in classification evaluation standards, dynamic adjustment mechanisms, and multi-subject participation, to explore the policy logic and implementation paths behind them. Second, research on the localization adaptability of international advanced evaluation models should be strengthened. For example, although Germany's "Professor Performance Agreement" and Australia's "Excellence in Research for Australia" have reference value, their successful experiences are closely related to specific national legal frameworks, academic traditions, and resource allocation models. Future research needs to use the case study method to benchmark countries with development levels similar to China's universities, deeply analyze the institutional obstacles and cultural conflicts that their evaluation systems might face during the transplantation process, and propose targeted adaptation strategies. Simultaneously, combined with the "Belt and Road" education cooperation background, distinctive teacher evaluation paths for developing countries can be explored to provide Chinese solutions for global higher education governance. Furthermore, international comparison of cross-disciplinary evaluation urgently needs to be strengthened. The international academic community is still in the exploratory stage regarding the evaluation of interdisciplinary teachers, and a unified standard has not yet been formed. Future research can focus on the practices of the world's top universities in interdisciplinary team performance assessment, achievement attribution, and collaborative innovation incentives, distilling quantifiable evaluation indicators and dynamic weight allocation models. By comparing and analyzing the policy support intensity and resource allocation efficiency of different countries in interdisciplinary evaluation, theoretical bases can be provided for Chinese universities to break down disciplinary barriers and promote cross-integration. Finally, a long-term tracking mechanism for international comparative research should be constructed. Teacher evaluation reform is a dynamically evolving process, and policies and practices in various countries are constantly adjusting. It is recommended to establish an international university teacher evaluation database to regularly collect and analyze policy texts, implementation cases, and effect assessment data from various countries, utilizing text mining and econometric analysis methods to reveal the development trends and internal laws of global teacher evaluation reform. This mechanism will not only enhance the timeliness and foresight of China's research but also provide data support for educational administrative departments to formulate policies with a more international vision.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Wang Zhanjun, Liu Jing, Qiao Gang. Clearing the "Four Onlys" Calls for Evaluation Innovation in "Double First-Class" Construction. China Higher Education, 2019(01): 16-19+26.
[2] Tian Xianpeng, Jiang Shujie. From the "Five Onlys" to "Multiple Dimensions": The Reform of Academic Evaluation for University Teachers. Shanghai Journal of Educational Evaluation, 2022, 11(03): 1-5.
[3] Cao Ting. Construction of a Comprehensive Quality Evaluation System for "Dual-Qualified" Teachers under the Background of "Breaking the Five Onlys". Modern Business Trade Industry, 2024, 45(07): 135-138.
[4] Liu Zhentian. From the "Five Onlys" and Multiple Dimensions to Self-Direction: Where Should the Reform of University Education Evaluation Break Through? University Education Science, 2024(5).
[5] Cao Taisheng. The Problem of the "Five Onlys": Consequences, Root Causes, and Solutions for Faculty Evaluation in Universities. University Education Science, 2019(1).
[6] Dong Yunchuan, Feng Yuanyuan. On the Contemporary Dilemmas of the Academic Community in Universities. Academic Exploration, 2024(5).
[7] Gu J X, Levin J. Tournament in Academia: A Comparative Analysis of Faculty Evaluation Systems in Research Universities in China and the USA. Higher Education, 2021, 81(5): 897-915.
[8] Thornton P H, Ocasio W. Institutional Logics and the Historical Contingency of Power in Organizations: Executive Succession in the Higher Education Publishing Industry, 1958-1990. American Journal of Sociology, 1999, 105(3): 801-843.

[9] Luo Yan. An Institutional Analysis of the "Five Onlys" Academic Evaluation — Concurrently Discussing the Institutional Orientation of China's Academic Evaluation After Opposing the "Five Onlys". Fudan Education Forum, 2020(3).

[10] Xiao Guofang, Peng Shulian. Rational Reflection and Reform Direction of Academic Evaluation in Universities under the Background of "Breaking the Five Onlys". Scientific Management Research, 2021(2): 65-72.

[11] Qin Yifan, Zhang Xinping. Path Dependence and Breakthrough of Faculty Evaluation System in China's Universities. Modern University Education, 2024(2): 86-92.

[12] Chen Chunli. Construction of an Evaluation Index System for "Four-Have" Good Teachers in Vocational Education under the Background of Chinese Modernization. Vocational and Technical Education, 2024, 45(05): 71-75.

[13] Jermolajeva J, Lchenkova S. Professional Roles as the Structural Component of Professional Identity of Higher Education Teachers in Samples of Riga and Smolensk. Economics and Culture, 2017, 14(2): 5-12.

[14] Luo Yun. After "Breaking the Five Onlys", How Can the Evaluation and Appointment of Professional Titles in Universities Become More Effective? Guangming Daily, 2021.

[15] Tu Duanwu. Policy Promotion, Problems, and Suggestions for Education Evaluation Reform — A "Dialogue" Between Policy Text and Practice. Fudan Education Forum, 2020, 18(02): 79-85.

[16] Ma R, Wang H W. Faculty Evaluation Reform Based on Multi-Case Study: Progress, Dilemma and Direction. Journal of National Academy of Education Administration, 2023(2): 52-61.

[17] Song Y H, Zhu L, Qiu J P. Reflections on the Construction of Scientific Research Evaluation System under the Background of "Breaking the Five Focuses". Journal of Intelligence, 2022, 41(2): 190-197.