



OPEN A faculty performance evaluation model based on MACBETH and fuzzy filter ranking methods

Min Tao¹, Quanxin Chen¹, Wenyan Yan¹ & Xiong Wang²✉

The establishment of a robust faculty performance evaluation system has become a research hotspot, as it is crucial for the continuous improvement and sustainable development of higher education institutions. However, according to available literature, the need for an easy-to-use management tool for faculty performance evaluation—one that accounts for professional tracks—within the context of China's application-oriented universities remains unaddressed. This study presents a new evaluation framework for measuring and ranking faculty performance, tailored to the characteristics of application-oriented universities. The proposed model applies hybrid multi-criteria decision making (MCDM) techniques through a three-step approach. First, a set of measurement indicators is developed based on a comprehensive review of existing literature and faculty assessment forms used by application-oriented universities in practice. Second, in line with the classification of faculty members' career tracks, the MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) method is utilized to determine the weights of evaluation criteria for each professional track. Finally, the FFR (Fuzzy Filter Ranking) approach is employed to rank faculty performance and identify outstanding faculty award winners. The applicability and utility of the proposed methodology are validated through a case study of an application-oriented university in China, which demonstrates its value as an effective evaluation tool and decision-aiding reference for faculty performance assessment. This study offers an intuitive and readily applicable solution for university stakeholders, streamlining the processes of weight assignment and faculty ranking within performance evaluations. Moreover, it can be extended to address other evaluation problems in academia, thereby contributing to the enhancement of educational quality.

Keywords Faculty performance evaluation, MACBETH, Fuzzy filter ranking, Higher education institutions

Faculty performance evaluation is a periodic and systematic process conducted within higher education institutions to measure and assess the effectiveness of faculty members in key aspects of their work, such as teaching, scholarship, and service. A proper evaluation of faculty performance generates mutual benefits for both faculty and institutions. Research has shown that an effective performance evaluation system can positively influence faculty members' work enthusiasm, teaching quality, and research output; enhance the overall standing of universities and colleges; and foster a culture of continuous improvement and excellence within higher education institutions¹. By identifying the strengths and weaknesses of its faculty members, an institution can also allocate resources more effectively, prioritize efforts in faculty recruitment and retention, and initiate appropriate faculty development programs.

Application-oriented universities in China have a distinct role in the higher education landscape. According to China's Ministry of Education, there were 672 application-oriented higher education institutions in 2022, accounting for 52.9% of the country's colleges and universities. These institutions primarily focus on undergraduate education, with an emphasis on practical skills and industry-relevant knowledge to prepare students for professional careers. Their goal is to cultivate talents who not only have a solid grasp of theoretical foundations but also possess practical skills and innovative abilities, enabling them to effectively meet the demands of industry and social development. Application-oriented universities differ from research-intensive universities in terms of educational objectives, teaching models, service areas, faculty capability requirements, and student development pathways. This makes them a valuable choice for students seeking a practical, career-oriented education. However, universities and colleges are currently dominated by research-based performance

¹Business School, Nanjing Xiaozhuang University, Nanjing, Jiangsu, China. ²BRU-IUL, ISCTE Business School, University Institute of Lisbon, Lisbon, Portugal. ✉email: wang_xiong@iscte-iul.pt

evaluation systems², which may be detrimental to creativity and innovation in teaching^{3,4}. Such systems are also not suitable for application-oriented universities.

How to effectively address faculty performance evaluation issues in the complex context of application-oriented universities is a challenging problem. In essence, faculty performance evaluation can be regarded as a complex multi-criteria decision making (MCDM) problem, involving multiple conflicting attributes, diverse stakeholder interests, and ambiguity. MCDM methods demonstrate significant potential in faculty performance evaluation, offering an effective approach to enhance the scientific rigor, comprehensiveness, and objectivity of the evaluation process. To promote the high-quality development of higher education and enhance the scientific decision-making of policymakers, this study proposes a technical tool for faculty performance evaluation that enables the precise expression and depiction of fuzzy and uncertain preference information in the evaluation process. This study adopts a combined method to assess faculty performance within the context of China's application-oriented higher education institutions. Specifically, the Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH) is applied to calculate the weights of performance evaluation criteria for different faculty career tracks. Subsequently, the Fuzzy Filter Ranking (FFR) approach is utilized to rank faculty performance based on the obtained weights, as well as individual ratings or actual achievements in each criterion. In this way, this study addresses the following research questions: (1) What constitutes an effective and transparent method for evaluating faculty performance? (2) What metrics and factors should be employed, considering the characteristics of application-oriented universities? (3) How should winners of the outstanding faculty award be selected?

The selection of the integrated MACBETH-FFR approach is justified as follows.

One of the key advantages of the MACBETH approach lies in its requirement for only qualitative judgments regarding the differences in relative attractiveness values between pairs of options. It can also handle situations where providing precise values is not feasible. In contrast, many other classic weighting methods in MCDM require the elicitation of numerical values, a task acknowledged as challenging⁵. Furthermore, the questioning mechanism of MACBETH is straightforward and intuitive. Each query in MACBETH involves only two options, facilitating a more tractable decision-making process. The approach also features a comprehensive procedure that is lacking in other techniques. For instance, it defines indicators associated with each criterion, ensures that criteria can be compared on a common scale, and maintains consistency in the judgments provided⁶. Additionally, the availability of M-MACBETH software streamlines the process by enabling decision makers to obtain results automatically.

Meanwhile, the FFR method offers several distinct advantages⁷. First, it functions as a more flexible tool for decision makers to manage ranking tasks, enabling manual adjustments to the filter threshold. Second, it reduces the difficulty of arithmetic calculations and the complexity of the ranking process, thereby significantly simplifying the practical application of MCDM methods. Third, it mitigates the influence of extreme alternatives on ranking outcomes. Fourth, it does not require verifying the consistency of estimates, while allowing comparisons even when criteria use different measurement units⁸. Additionally, this method can be easily programmed into a computer application to automatically rank alternatives, making it particularly suitable for handling large-scale ranking tasks involving numerous alternatives.

The paper makes the following threefold contributions:

First, this study proposes a practical methodology that avoids complex mathematical operations. It presents a multi-criteria decision support system for evaluating faculty performance and identifying outstanding faculty members through ranking-based solutions. The integration of the MACBETH and FFR methods ensures user-friendliness for university stakeholders. Bernroider and Schmöller⁹ surveyed Austrian companies and collected 114 completed questionnaires, finding that 71.9% of the companies were aware of MCDM methods, but only 33.3% actually used them in practice. Štilić and Puška¹⁰ noted that presenting intuitive processes, visualizations, and case study-based explanations is an effective way to promote the adoption of useful MCDM methods. This paper thus contributes to bridging the identified gap between known and applied MCDM methods by exploring user-friendly decision support tools.

Second, this paper promotes interdisciplinary research in education by adopting a design science research process and demonstrates the application of MCDM methods in the educational field. Given the growing significance of decision-making in contemporary contexts, scholars have expressed a need for more MCDM techniques and approaches to be applied in higher education¹¹. Yüksel et al.¹² have also called for studies that employ MCDM techniques distinct from those used in existing research. This study integrates technical and managerial knowledge to develop a more practical and useful solution methodology. Based on existing literature, this paper represents the first application of the combined MACBETH and FFR methods in higher education.

Third, this research contributes to the literature on performance management by developing a multi-criteria evaluation index and extends educational research within the context of application-oriented universities in China. Notably, the evaluation framework incorporates criteria unique to Chinese universities (e.g., political and ideological performance). The study presents rankings for three faculty categories—teaching-focused, teaching-research balanced, and research-focused faculty—a distinction that has not been addressed in previous literature. It also serves as a reference for practitioners and software companies. With the proposed framework, each academic department can adopt simple quantitative methods to develop an evaluation system aligned with its objectives. Furthermore, human resource software companies can integrate the proposed model into the performance management module of their products to enhance functionality.

The remainder of the paper is structured as follows. [Literature review](#) presents a literature review of performance measurement models and criteria in higher education, as well as the application of MCDM methods in faculty performance evaluation, followed by the methodological background in [Methods](#). [The proposed framework](#) describes the proposed framework with detailed procedural steps. [Case study](#) includes a case study to implement the proposed methodology. Additionally, managerial insights are derived from the

analysis and discussion of the results. Finally, [Concluding remarks](#) presents the conclusions and suggestions for future research.

Literature review
MCDM applications in higher education evaluation

MCDM techniques are primarily used to evaluate the efficiency and performance of higher education institutions and to inform decision-making processes. Systematic literature reviews on the application of MCDM techniques in higher education can be found in^{11–17}.

Academic performance evaluation covers areas such as accreditation and quality assurance, institutional performance evaluation, university rankings, student performance evaluation, academic staff performance evaluation, and curriculum design and assessment. Table 1 presents representative references related to the application of MCDM in academic performance evaluation, with a focus on weight assignment and subsequent result evaluation.

Evaluation criteria are formulated based on the evaluation purpose and specific application contexts (e.g., country-specific contexts). Due to the limitations of stand-alone methods, the adoption of hybrid MCDM methods has become a trend in addressing academic performance problems. Among MCDM practices in higher education, traditional weighting and ranking methods are most prevalent, including AHP (Analytic Hierarchy Process), DEA (Data Envelopment Analysis), VIKOR (VlseKriterijumska Optimizacija I Kompromisno Resenje), and TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution). In terms of illustrative examples of MCDM models, previous studies have primarily focused on research-oriented universities.

MCDM applications in faculty performance evaluation

An effective and robust faculty performance evaluation system is of importance to universities as it serves as a tool for measuring educational quality, identifying areas for improvement, and recognizing excellence. Consequently, there is a growing focus on measuring the performance of faculty members within higher education institutions. To enhance the effectiveness of evaluation, it is necessary to improve the transparency and quality of the evaluation process, increase faculty members’ recognition of and satisfaction with the results, and link evaluation results with their promotions, awards, and bonuses. Traditional performance measurement

References	Research objective	Method	Criteria	Application
18	Assess faculty performance and rank faculty members within each discipline or major	AHP	Research, teaching and service	United Arab Emirates University
19	Assess the management performance of research and development (R&D) activities	AHP and DEA	Full-time R&D personnel, R&D expense, fruits of science and technology, international communications, and awards	29 research-oriented universities in China
20	University ranking	DEA	Total research; federal research; national academy members; faculty awards; doctorates granted; postdoctoral appointees; median SAT scores; endowment assets; annual giving; national merit and achievement scholars	Top American research universities
21	Evaluate the instructors’ performance in administration sciences in universities.	AHP, fuzzy set theory, and TOPSIS	Teaching style, individual features and social relation, knowledge level, observance of educational regulations, and educational tools	Islamic Azad University
22	University ranking	AHP and VIKOR	Administration and professional	12 private universities in Taiwan
23	Performance evaluation in higher education institutes	AHP and BSC	Financial, customer, inner process, and progress and learning	4 humanities colleges in Iran
24	Prioritize performance indicators in engineering education	AHP, TOPSIS and BSC	Financial, customer, inner process, learning and growth, and management commitments	Fayoum University (Egypt)
25	Compare tutors on teaching performance	AHP and DEA	Interest, usefulness, professionalism, preparation, presence, and supporting material	A higher education institution in Greece
26	Build composite indicators to evaluate university performance	Double reference point method	Research, teaching and technology transfer	9 public universities in Andalusía
27	Measure and evaluate the academic performance of students	uwTOPSIS	Academic, cognitive, economic, health, and social	Industrial University of Santander (Colombia)
28	Evaluate scientific research efficiency of faculty members	Modified meta frontier DEA	Journal publications and research grants	Inner Mongolia University (China)
29	Evaluate and rank the faculty of the Engineering School based on teaching performance	DEA	Total number of courses sections taught in the year; total number of different courses taught in the year; total number of students in these courses; total number of teaching hours in the year; weighted average evaluation of the quality of the teaching given by the students	Universidad de Antioquia– UdeA (Colombia)
30	Develop a university performance measurement framework	Fuzzy Delphi and BWM	Education, research, commercialization, human capital, educational facilities and infrastructures, social, cultural and welfare facilities, and popularity	A public university in Iran
31	University ranking	IVN-AHP and VIKOR	Learning experience, campus life, academic support, management, learning opportunities and resources, and personal development and career support	191 universities in Turkey
32	Academic performance ranking of research universities	Entropy and PROMETHEE	Article, citation, scientific document, doctorate score, faculty member / student, international cooperation, national cooperation, and TUBITAK Project	20 research universities in Turkey

Table 1. A summary of MCDM applications in higher education evaluation.

tools, such as the balanced scorecard and 360 feedback, have their limitations. MCDM methods, due to their ability to comprehensively consider multiple dimensions of factors, have received extensive attention in the field of faculty performance evaluation in recent years.

The AHP and TOPSIS methods, along with their fuzzy variants, are most widely used by scholars in constructing frameworks for faculty performance evaluation across various countries. For instance, Do³³ proposed a method combining fuzzy AHP and fuzzy TOPSIS to assess and rank lecturers' performance in the Vietnamese context. However, based on a comprehensive literature review, only 3 main criteria (i.e., teaching, research, and service) and 9 sub-criteria (e.g., quality of teaching content, success in securing external funding to support research, and service to the community) were identified for the evaluation, which is insufficient to assist evaluators in reviewing and measuring overall performance in detail. Karmaker et al.³⁴ introduced AHP and TOPSIS methods to evaluate the performance of seven teaching staff in an academic institution in Bangladesh. The evaluation criteria included subject knowledge, communication, discipline, cooperation, and creativity, with 3 sub-criteria under each main criterion. To evaluate teacher performance in an Indonesian vocational school across 4 competencies (pedagogic, personality, social, and professional). Hafizah et al.³⁵ designed a decision support system based on AHP and TOPSIS. Do et al.³⁶ proposed a new dynamic fuzzy TOPSIS method and applied it to a Vietnamese university without specifying the criteria.

The AHP method has also been integrated with other MCDM methods to evaluate teaching performance. Thanassoulis et al.²⁵ combined AHP and DEA to compare tutors' teaching performance, illustrating the methods with a case study of a higher education institution in Greece. Daniawan³⁷ used AHP and SAW (Simple Additive Weighting) to rank 28 lecturers at an Indonesian university based on 10 criteria of teaching quality (e.g., the updating and relevance of content to current conditions, overall assessment, and clarity in content delivery and example provision). Although both the AHP and MACBETH approaches employ pairwise comparisons, the scales used by decision makers to make judgments differ. AHP utilizes a 9-point ratio scale, whereas MACBETH adopts an ordinal scale with seven semantic values. For weight determination, AHP employs an eigenvalue method, while MACBETH relies on linear programming. In each matrix, AHP tolerates up to 10% inconsistency in judgments, whereas MACBETH does not permit any form of inconsistency⁶.

Many other scholars have also employed other MCDM methods with more complex mathematical models to address faculty performance evaluation problems. For instance, Radovanović et al.³⁸ implemented a hybrid MCDM model combining the Spherical Fuzzy AHP and Grey MARCOS (Measurement of Alternatives and Ranking According to Compromise Solution) methods to evaluate university professors based on seven criteria, including teaching quality, professor availability to students, student assessment, contribution to the university, competence, and ethics and professionalism. Aazagreyir et al.³⁹ integrated two distinct methodologies, namely Fuzzy DEMATEL (Decision Making Trial and Evaluation Laboratory) and SAW, to evaluate lecturers' performance based on the criteria of teaching, research, service, and commercialization. Watrionthos et al.⁴⁰ applied the PROMETHEE (Preference Ranking Organization Method for Enrichment Assessment) and GAIA (Geometric Analysis for Interactive Aid) methods to assess lecturer performance based on 5 metrics that focus solely on various types of publications. Complex programming models and calculation processes may make it difficult for users to understand and interpret the results, particularly for those without a strong background in technology, mathematics, or statistics. Agrell and Steuer⁴¹ proposed a combined DEA and linear programming method for the performance review of individual faculty, considering research output, teaching output, external service, internal service, and cost as criteria. The proposed multi-criteria decision support system was applied to a university department where 30 faculty members were evaluated over a 3-year period. Qian et al.⁴² combined the Triangular Fuzzy DEMATEL with the Backpropagation Neural Network (BP) to construct and optimize a teacher evaluation index system comprising 20 indicators (e.g., classroom teacher-student interaction, enthusiasm for one's own work), though its applicability and accuracy were not tested.

To summarize, there is a lack of research on comprehensive faculty performance evaluation across multiple dimensions in the context of application-oriented universities. Previous studies also apply a uniform set of measurement and evaluation weights to all faculty, ignoring the need to differentiate faculty roles (e.g., research faculty, clinical faculty) to ensure fair performance comparisons. In general, faculty members affiliated with a specific academic department have their own professional tracks. If the departmental assessment policy is not suitable for faculty engaged solely in teaching activities, their performance may be underestimated. Complicated models may also impede the application of decision support tools. To address these issues, this paper proposes an effective faculty performance evaluation system that employs integrated MCDM techniques to resolve problems related to criterion selection, weight determination, and data processing. This paper uses the performance of faculty members at a public application-oriented university as an example to facilitate a better understanding of the proposed model. The model aims to provide a fair and balanced performance assessment method that reflects actual performance, offering an easy-to-use, versatile, and objective methodology for prioritizing faculty members within an academic department for excellence awards, merit pay, tenure decisions, or other recognition schemes.

Methods

MACBETH

The MACBETH method has emerged as a prominent method for generating weights in the MCDM field. It was first proposed by e Costa and Vansnick⁴³. Due to its humanistic, interactive, and constructive nature, it has attracted the attention of scholars in the decision-making research community since its inception. Ferreira and Santos⁴⁴ conducted a bibliometric literature review on the application of the MACBETH technique.

When applying the MACBETH technique, decision makers (either an individual or a group) are invited to conduct pairwise comparisons among the options, actions, or alternatives in set A (where $A = \{a, b, \dots, n\}$ represents a finite set of n alternatives). They are asked to make semantic judgments about the difference in

attractiveness between these alternatives. The pairwise comparison is carried out according to 7 pre-defined semantic categories for the difference in attractiveness shown in Table 2. For example, if alternative a is considered more attractive than alternative b , and the difference between them is considered *strong*, then $(a, b) \in E4$. In contrast to some other methods like direct rating or bisection, where the decision maker must deal with significant cognitive effort due to the need to compare more than two alternatives, the MACBETH process involves only two alternatives at a time.

Through linear programming, the judgments of the decision maker derived from pairwise comparisons are converted into a MACBETH scale. Let $v(x)$ represent the score given to option X , where x^+ is at least as attractive as another option within X and x^- is at most equally as attractive as another option within X ⁶. The linear programming utilized is:

$$\text{Min } [v(x^+) - v(x^-)]$$

Subject to $v(x^-) = 0$ (arbitrarily assigned)

$$\forall (x, y) \in C_0 : v(x) - v(y) = 0$$

$$\forall (x, y) \in C_i \cup \dots \cup C_s \text{ with } i, s \in \{1, 2, 3, 4, 5, 6\} \text{ and } i \leq s : v(x) - v(y) \geq i$$

$$\forall (x, y) \in C_i \cup \dots \cup C_s \text{ and } \forall (w, z) \in i, s \in C_{i'} \cup \dots \cup C_{s'}$$

$$\text{with } i, s, i', s' \in \{1, 2, 3, 4, 5, 6\} \ i \leq s, i' \leq s', \text{ and } i > s' : v(x) - v(y) \geq v(w) - v(z) + i - s'$$

Unlike other multicriteria methods, MACBETH does not allow any inconsistency. The method features an automated process that can identify inconsistencies, even in scenarios where arrays of value judgments are incomplete. When this linear program turns out to be unfeasible, the judgments are regarded as inconsistent. Conversely, if it is feasible, there may be multiple optimal solutions. After achieving consistent value judgments, linear programming techniques can formulate cardinal value functions, which enable the calculation of trade-offs among variables⁴⁴. In this case, the mean is used as the MACBETH scale. This methodology allows for the generation of numerical scores for different alternatives and the determination of numerical values for criteria weights.

MACBETH also requires a descriptor to be defined for each criterion to be evaluated. A descriptor is an ordered set of possible impact levels associated with a criterion, whose purpose is to objectively depict the influence of the alternatives with respect to that specific criterion. The higher the degree of objectivity in creating the descriptor, the less ambiguity there will be, and the easier it will be for the model to be understood and accepted. Two of the descriptor's levels are used as reference points. One is named "Good", which the decision maker deems unquestionably satisfactory. The other is "Neutral", representing a level that the decision maker considers neither satisfactory nor unsatisfactory. Experiments have verified that this approach notably enhances the intelligibility of the criterion.

Fuzzy filter ranking

The FFR method is a compromising method proposed by Chang and Ku⁷ to determine the ranking of alternatives. The basic concept of the FFR method is to combine the filtering method, discrete fuzzy score, and Likert scale. Based on the weights obtained from the MACBETH approach and the values of each criterion, faculty performance can be ranked following the FFR steps. This method offers a simple and flexible means of producing reliable ranking results. The calculation steps of the FFR method are as follows.

Step 1: Criterion value classification and fuzzy score mapping. Based on many random ranking experiments, Chang and Ku⁷ have demonstrated that a 5-point Likert scale seems to be the best choice for the scale of filtering. Hence, all criterion values for the alternatives are categorized into 5 groups, which are mapped to the discrete fuzzy score range [0, 1]. Specifically: (1) a strongly significant criterion corresponds to a fuzzy score of 1; (2) a significant criterion corresponds to a fuzzy score of 0.75; (3) a normal criterion corresponds to a fuzzy score of 0.5; (4) a non-significant criterion corresponds to a fuzzy score of 0.25; and (5) a strongly non-significant criterion corresponds to a fuzzy score of 0.

Step 2: Construction of filters. Four filters are constructed as boundaries for the above five scales, forming five regions. Filter 1, denoted as A_i^{LI} , serves as the threshold distinguishing between strongly non-significant and

Code	Semantic scale	Description
E0	No difference	Indifference between the alternatives
E1	Very weak	a is very weakly attractive over b
E2	Weak	a is weakly attractive over b
E3	Moderate	a is moderately attractive over b
E4	Strong	a is strongly attractive over b
E5	Very strong	a is very strongly attractive over b
E6	Extreme	a is extremely attractive over b

Table 2. MACBETH semantic scale.

non-significant criteria. Similarly defined are A_i^{Lh} , A_i^{Ul} , and A_i^{Uh} , where $A_i^{Ll} < A_i^{Lh} < A_i^{Ul} < A_i^{Uh}$. A_i^{Lh} acts as the boundary between non-significant and normal criteria, A_i^{Ul} separates normal and significant criteria, and A_i^{Uh} provides a boundary between significant and strongly significant criteria. Given the difficulty in analyzing the criterion dataset to derive the probability distribution, a simple method—outlined in Eq. (1) to (4)—is proposed by Chang and Ku⁷ to obtain the respective filter for the i th criterion. In this method, the average value of the criteria is set as the center, and then the boundaries are deployed on both sides. The lower area [A_i^{min} , A_i^{avg}] is divided into three subareas using $A_i^{min} - A_i^{Ll} = A_i^{Ll} - A_i^{Lh}$ and $A_i^{Lh} - A_i^{avg} = 0.5(A_i^{Ll} - A_i^{Lh})$. The upper area [A_i^{avg} , A_i^{max}] is divided into three subareas using $A_i^{max} - A_i^{Uh} = A_i^{Uh} - A_i^{Ul}$ and $A_i^{Ul} - A_i^{avg} = 0.5(A_i^{Uh} - A_i^{Ul})$. The notations A_i^{avg} , A_i^{max} , and A_i^{min} represent the average value, maximum value, and minimum value of the alternatives in the i th criterion, respectively.

$$A_i^{Ll} = \frac{3(A_i^{min} - A_i^{avg})}{5} + A_i^{avg} \quad (1)$$

$$A_i^{Lh} = \frac{1(A_i^{min} - A_i^{avg})}{5} + A_i^{avg} \quad (2)$$

$$A_i^{Ul} = \frac{1(A_i^{max} - A_i^{avg})}{5} + A_i^{avg} \quad (3)$$

$$A_i^{Uh} = \frac{3(A_i^{max} - A_i^{avg})}{5} + A_i^{avg} \quad (4)$$

Step 3: Creation of a transition decision matrix. A transition decision matrix, m_{ij} (transition value of the i th criterion in the j th alternative), is created. By quantifying each region with constant values of 1, 0.75, 0.5, 0.25, and 0, this matrix reflects the discontinuous perception of decision makers in most MCDM problems. Specifically, if the value of the i th criterion for the j th alternative is less than the first filter of the i th criterion, it is mapped to 0; if the value lies between the first and second filters, the output value of the transition decision matrix m_{ij} equals 0.25; likewise, if the value lies between the second and third filters, it is mapped to 0.5; if the value lies between the third and fourth filters, it is mapped to 0.75; finally, if the input value is above the fourth filter, the output value equals 1.

Step 4: Calculation of weight correctors. The weight correctors are calculated using Eq. (5). Since the output values of the transition decision matrix m_{ij} may not be equal across criteria, and this inequality affects the weight deviation of the criteria, a correction value is used to ensure that the sum of the output values for each criterion equals 1.

$$w_i^m = 1 / \sum_{j=1}^J m_{ij} \quad (5)$$

Step 5: Setting the effect factor. The effect factor f_i is set to 1 for benefit criteria and -1 for cost criteria.

Step 6: Generation of the weighted and corrected decision matrix s_{ij} . According to Eq. (6), it is obtained by multiplying the weight w_i by the effect factor f_i , the weight corrector w_i^m , and the transition decision matrix m_{ij} , where w_i is the weight of the i th criterion derived from the MACBETH approach in this study, and $\sum_{i=1}^I w_i = 1$. Finally, the ranking of the alternatives is determined according to the sum of s_{ij} for each alternative.

$$s_{ij} = w_i \times w_i^m \times f_i \times m_{ij} \quad (6)$$

The proposed framework

Faculty performance evaluation is a complex and multifaceted endeavor that demands meticulous consideration of diverse factors and stakeholders. It can be viewed as a systematic process, as illustrated in Fig. 1. The inputs to the entire process are a list of faculty members and their job responsibilities, taking into account individual strengths and weaknesses. Based on these inputs, each faculty member can be classified into an applicable group (or career track). The core of this system lies in obtaining the actual performance results of each faculty member and the performance evaluation criteria. Information sources include university and college strategies (e.g., annual objectives) and administrative software (e.g., lists of publications, student satisfaction surveys, registered service hours). All faculty members within the same group are then compared and ranked. The final output is a proposal for outstanding faculty award winners from each group.

To assist decision makers in making more informed decisions during the faculty performance evaluation process, we propose a multi-phase framework that integrates multiple methods, as depicted in Fig. 2. This framework is designed around the essential elements of a comprehensive evaluation and consists of three primary steps: (1) identifying and selecting evaluation criteria based on a literature review and real-world evaluation forms collected from application-oriented universities; (2) determining the criteria weights by categories of faculty career tracks using the MACBETH approach; and (3) assessing faculty performance with the FFR method to establish the ultimate ranking. The following subsections elaborate on the implementation steps of the illustrated framework, aiming to provide a general understanding of the proposed faculty performance evaluation process.

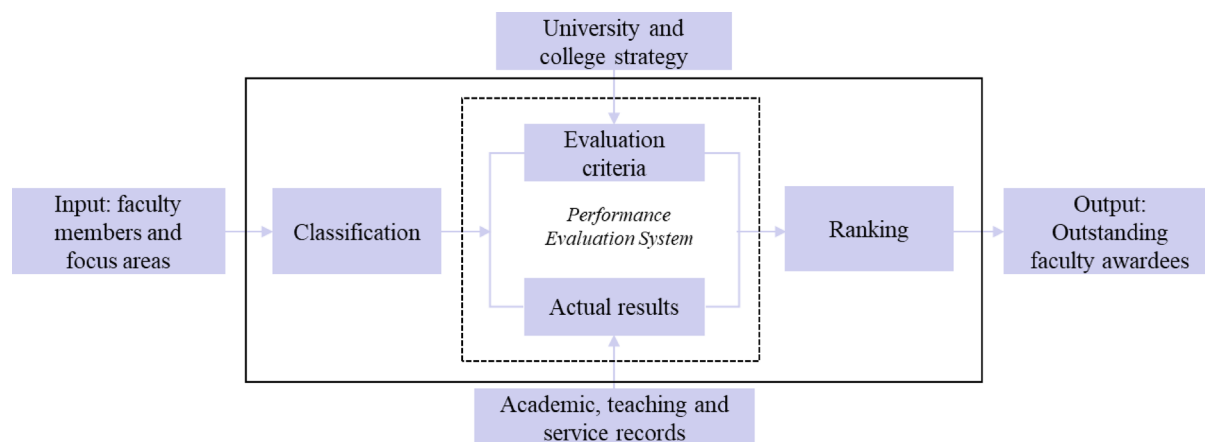


Fig. 1. A systematic faculty performance evaluation process.

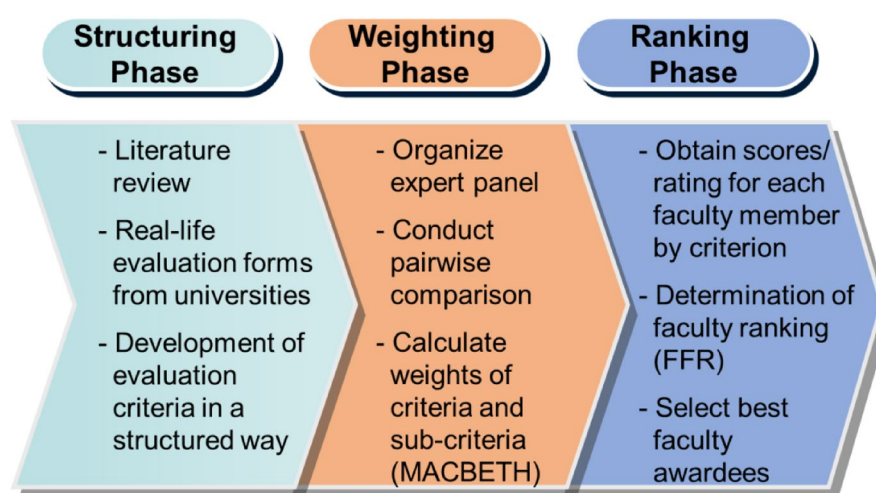


Fig. 2. The proposed framework for faculty performance evaluation.

Determination of evaluation criteria

The determination of assessment metrics is of great significance in the evaluation of faculty performance. Evaluation indicators and metrics used in faculty performance evaluations vary widely across higher education institutions. In general, the evaluation framework should encompass diverse aspects such as teaching, research, social contributions, and professional ethics. Meanwhile, it should align with the university's strategic positioning and characteristics, ensuring coherence, scientific rigor, and comprehensiveness.

Drawing on the criteria from the prior research reviewed in Sect. 2 and referencing other faculty performance evaluation models^{45–47}, we synthesized general faculty performance evaluation criteria for higher education institutions. In addition, we collected evaluation forms from 16 application-oriented universities in China, including the case-study university, to identify the most frequently used evaluation elements unique to application-oriented universities in China (e.g., political and ideological performance). Ultimately, we developed an evaluation framework consisting of 24 individual criteria grouped into 5 clusters, designated as C_{11} , C_{12} , ..., C_{mn} , as shown in Table 3. Given the nature of application-oriented universities, it is not surprising that they place significant emphasis on teaching-related activities, which is reflected in the greater number of dimensions in this area.

Weight generation through MACBETH

The evaluation is extremely sensitive to the selection of weights. Prior research has demonstrated that in numerous instances, even slight fluctuations in weights can lead to different rankings of faculty performance⁴⁸. Moreover, the duties and responsibilities of faculty members in different career tracks vary significantly, and thus the focus areas of evaluation must be differentiated accordingly. Therefore, different weights should be determined for each criterion based on the classification of career tracks proposed in Table 4.

When using the MACBETH approach to calculate weights, the initial step is to formulate value functions for each criterion. The decision maker provides judgments between the scale levels of each descriptor using

Main criteria	Sub-criteria	Description	Measurement	Remarks
Professional ethics and conduct (C1)	Political and ideological performance (C11)	Follow the Party Central Committee closely in terms of political stance, direction, principle, and path.	Qualitative	Dean and academic leaders of the college evaluate each faculty member according to the following scale: 1 = Below expectations; 2 = Meets expectations; 3 = Exceeds expectations; 4 = Greatly exceeds expectations; 5 = Truly outstanding.
	Professional ethics and literacy (C12)	Have a good work ethic, be strict in self-discipline and a role model for students.	Qualitative	
	Teamwork awareness (C13)	Become a good team player, actively participate in departmental and institutional activities and social responsibility activities.	Qualitative	
Service and leadership (C2)	Service work (C21)	Participate in university service, such as committee and leadership roles	Qualitative	The same scale as C1
	Coach junior lecturer (C22)	Help in recruitment and mentor new faculty members	Qualitative	
	Industry engagement (C23)	Connect and collaborate with industry to improve practical skills	Qualitative	
Teaching activities (C3)	Teaching workload (C31)	The teaching hours completed per year (including thesis and internship supervision).	Quantitative	N/A
	Student evaluation (C32)	The score from the student evaluation on teaching.	Quantitative	
	Peer evaluation (C33)	The score from the peer evaluation on teaching.	Quantitative	
	International student class (C34)	The number of courses taught to international students of the college.	Quantitative	Cost criteria
	Teaching & learning rules (C35)	The number of cases for late arrival or early departure of the class, change of the finalized class schedules or absence from required meetings (e.g., faculty training) without pre-approval and proper justification, failure to submit required materials and file test paper timely, etc.	Quantitative	
Teaching achievements (C4)	Teaching achievement awards (C41)	The number of awards based on the level (e.g., national, provincial, school) and rank (e.g., first, second, third).	Quantitative	Defined scale for each award: (1) National award: 100; (2) Provincial award: 50 = grand prize; 40 = first prize; 25 = second prize; 20 = third prize; (3) School award: 15 = grand prize; 10 = first prize; 8 = second prize; 6 = third prize.
	Teaching reform projects (C42)	The number of approved projects on teaching reform based on level (e.g., national, provincial, school) and category (e.g., major, general).	Quantitative	Defined scale for each project: (1) Provincial level: 15; (2) Education Ministry's collaborative education program with industry: 12; (3) School level: 5 = major project; 3 = general project.
	First-class curriculum development (C43)	The number of first-class curricula based on level (e.g., national, provincial, school).	Quantitative	Defined scale for each curriculum: (1) National level: 40; (2) Provincial level: 20; (3) School level: 5 = major project; 3 = general project.
	Publication of textbooks and articles in teaching (C44)	The number of published textbooks based on category (e.g., major, general) and the number of published papers on teaching topics.	Quantitative	Defined scale for each textbook/paper (1) Provincial key textbook: 15; (2) Provincial general textbook: 6; (3) Articles on teaching topics: 2.
	Mentoring provincial college students' innovation program (C45)	The number of approved programs mentored by corresponding faculty members based on program category (e.g., major, general).	Quantitative	Defined scale for each program: (1) Major program: 8; (2) General program: 6; (3) Instructional program: 4.
	Teaching skill competitions (C46)	The number of awards based on the level (e.g., provincial, school) and rank (e.g., first, second, third).	Quantitative	Defined scale for each award: (1) Provincial award: 10 = first prize; 8 = second prize; 6 = third prize; (3) School award: 6 = first prize; 4 = second prize; 2 = third prize.
	Instruct students in various competitions (C47)	The number of awards won by the instructed students based on the competition category, level (e.g., national, provincial) and rank (e.g., first, second, third).	Quantitative	Defined scale for each award of Category I competition: (1) National award: 40 = first prize; 30 = second prize; 20 = third prize; (2) Provincial award: 12 = first prize; 8 = second prize; 6 = third prize. Defined scale for each award of Category II competition: (1) National award: 10 = first prize; 8 = second prize; 6 = third prize; (2) Provincial award: 6 = first prize; 4 = second prize; 2 = third prize.
	Outstanding graduation thesis awards (C48)	The number of supervised students who win outstanding graduation thesis award based on the rank (e.g., first, second, third).	Quantitative	Defined scale for each provincial award: 10 = first prize; 6 = second prize; 4 = third prize
	Laboratory construction (C49)	The total approved amount (capital spending) for constructing new labs.	Quantitative	The project leader shall propose proportion of the project amount for the participants based on contribution to the lab construction.
Continued				

Main criteria	Sub-criteria	Description	Measurement	Remarks
Research and scholarly activities (C5)	Publication records and patents (C51)	The number of approved patents and published articles and books based on the class/level of the journal, publisher, and patent.	Quantitative	Defined scale for each published paper: 5 = Class IV; 8 = Class III; 10 = Class II; 15 = Class I. Defined scale for each book: 10 = general press; 15 = major press. Defined scale for each patent: 10 = national invention patent, 2 = utility-model patent; 2 = registered software copyright.
	Scientific research projects (C52)	The number of approved projects based on the level of sponsors (e.g., national, provincial, municipal)	Quantitative	Defined scale for each project: (1) National level: 40; (2) Provincial level: 20; (3) Municipal: 5.
	Scientific research awards (C53)	The number of awards based on the level (e.g., national, provincial, municipal) and rank (e.g., first, second, third).	Quantitative	Defined scale for each award: (1) National award: 100; (2) Provincial award: 50 = first prize; 40 = second prize; 30 = third prize; (3) Municipal award: 15 = first prize; 10 = second prize; 7 = third prize.
	Horizontal scientific research projects (C54)	The amount of received research funding from enterprises and other organizations.	Quantitative	N/A

Table 3. The proposed evaluation criteria.

Type	Description
Teaching faculty	These faculty members are primarily engaged in teaching and contributing to the university’s curricular development. They are responsible for designing course syllabi, preparing teaching materials, implementing diverse teaching methods, and assessing students’ learning. They spend most of their time monitoring students’ learning progress and improving teaching quality through curriculum construction and teaching reform. They usually undertake little research work, which is mainly related to teaching methods and curriculum design.
Research faculty	These faculty members are expected to focus on conducting research, either independently or in collaboration with other faculty members. They are responsible for designing and leading research projects, securing research funds, and publishing academic papers. They usually have relatively few teaching tasks, mainly guiding students’ research or teaching specialized courses related to their research.
Teaching-research balanced faculty	These faculty members need to maintain a relatively balanced development in both research and teaching. They conduct research to a certain level, strive for research projects, and promote the transformation of achievements. Meanwhile, they undertake an appropriate amount of teaching tasks, actively participate in teaching reform, and integrate research results into teaching.

Table 4. The types of faculty members based on career track.

the MACBETH semantic scale. When it is impossible to precisely determine the difference in attractiveness between scale levels, a positive category can be utilized. This characteristic of MACBETH is highly beneficial for representing the decision maker’s uncertainty when making judgments. The MACBETH questioning process should be employed to populate the judgment matrix. First, compare the most attractive level of each descriptor with the least attractive one. Subsequently, compare the second most attractive level with the least attractive level, and so forth. Next, compare the most attractive level with the remaining options in descending order of attractiveness. Then, finalize the judgments that form the diagonal boundary of the upper triangular part of the matrix. Finally, provide the remaining judgments from the upper diagonal. Through linear programming, the M-MACBETH software generates a value function that assigns a value of 100 to the “Good” level and a value of 0 to the “Neutral” level. The decision maker should review the resulting value functions to ensure they accurately reflect the relative magnitude of their judgments. To complete the weighting process across criteria, an additional alternative must be created. This alternative should include all criteria at the “Neutral” level in every descriptor. The decision-maker should then use the MACBETH semantic categories to make judgments that assess the increase in attractiveness resulting from a shift from the “Neutral” level to the “Good” level in one of the descriptors. This step enables the criteria to be ranked from the most attractive to the least attractive. Following this, compare the most attractive swing with the second most attractive swing, then the most attractive swing with the third most attractive swing, and so on. This process continues row by row until the matrix is fully completed.

Based on the judgments, the M-MACBETH software calculates the weightings for each criterion and presents the values as percentages. A set of weighting values aligns with the decision maker’s judgements. These values form ranges that serve as the thresholds (maximum and minimum values) of the fuzzy numbers. The value assigned and verified by the decision maker is adopted as the modal value. The M-MACBETH decision support system and its user guide can be accessed at <https://www.m-macbeth.com>.

Faculty ranking through FFR

The FFR method selects the average value as the center of the criterion and then divides the upper and lower areas into subareas. Let us take criterion C_1 with five alternatives (A_1, A_2, \dots, A_5) as an example. These five alternatives are then placed into a funnel-shaped model featuring four filters. The boundaries of these filters are determined according to Eq. (1) to (4). Due to the distinct thresholds of the filters, these alternatives may stop at different layers, as shown in Fig. 3. In fact, for different criteria, different funnel-shaped models can be designed to meet diverse requirements. If the probability distribution (such as the normal or uniform distribution) of the criteria is known beforehand, the decision maker can select suitable thresholds for the filters. This enables the generation of a more precise ranking of the alternatives.

Based on the position of alternatives in all criteria and corresponding criteria weights, an overall score can be obtained and alternatives are ranked accordingly.

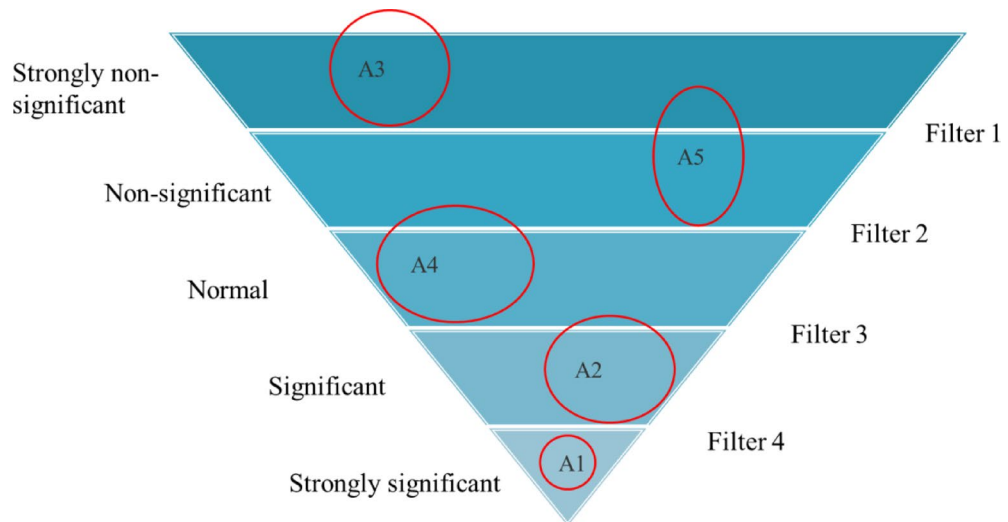


Fig. 3. Funnel-shaped model (reproduced from Chang and Ku⁷).

Case study

To demonstrate the implementation procedure of the proposed framework, a real-life case is presented to evaluate faculty performance at College X, which is part of an application-oriented university in China. The university has a long-standing history, tracing back to a school founded in 1927. As a public institution, it focuses on undergraduate education, with only a small number of graduate programs. It comprises 15 academic colleges and offers 47 undergraduate majors, catering to 18,000 full-time undergraduate students, while jointly cultivating over 130 graduate students with other research-oriented universities. The university employs 1,450 faculty members. College X, the subject of this study, is the second largest college in the university in terms of both faculty and student numbers. Currently, it has 61 full-time academic staff, approximately 80% of whom hold doctoral degrees, and offers 5 undergraduate majors with around 1,400 enrolled students. The current evaluation process at College X simply calculates a total score for each faculty member based on pre-defined criteria and corresponding scales. Outstanding faculty awards are then granted to the top 10 faculty members in the college. However, this process fails to account for the relative importance of different criteria through weight assignment and does not categorize faculty members based on their characteristics and responsibilities. Consequently, concerns regarding fairness and comparability have arisen.

For this case study, an online focus group was formed as the expert panel, consisting of three members (1 dean and 2 associate deans). The decision-making group reviewed and validated the criteria identified in Table 3 and their associated scales. Given that all key evaluation items currently used by the college were already included, no modifications were proposed. In the following subsections, the proposed set of criteria will be applied to construct a multi-criteria evaluation framework that reflects the priorities of the studied college.

Weighting the performance evaluation indicators

To determine the weights of the criteria, we first requested the decision-making group to categorize all faculty members into one of the three career tracks outlined in Table 4, considering their characteristics and responsibilities. Among the 61 faculty members at College X, 57% were identified as teaching-research balanced faculty, 33% were classified as teaching faculty, and the remaining percentage comprised research faculty. The following example demonstrates the MACBETH-based process for generating criteria weights to evaluate teaching faculty.

The first step is to rank the criteria in terms of their overall attractiveness. The 24 proposed evaluation indicators were ranked from the most important to the least important, with the aim of improving the performance of teaching faculty. After ordering all the performance evaluation criteria, the M-MACBETH weighting judgment matrix is to be filled out. The decision makers made qualitative judgments to assess the difference in attractiveness between the neutral and good levels of performance for different indicators, using the MACBETH scale. The results are presented in Figure A1 of the Supplementary Figures.

During the evaluation process, when the decision makers held divergent opinions, they engaged in constructive communication and exchanged ideas until reaching a consensus. This approach streamlined the decision-making process and facilitated agreement. The M-MACBETH software conducted an automated consistency check on the matrix. In the event of detected issues, it provided modification suggestions. Upon completion of the consistency check, the software generated an initial weighting scale proposal, which was then verified by the expert group. Decision makers could adjust the weight of a criterion within an interval generated by the software while ensuring it remained compatible with the judgment weighting matrix. The final weights are presented in Fig. 4. From a management perspective, the top three criteria for evaluating teaching faculty performance are student evaluation (C_{32}), first-class curriculum development (C_{43}), and teaching skill competitions (C_{46}).

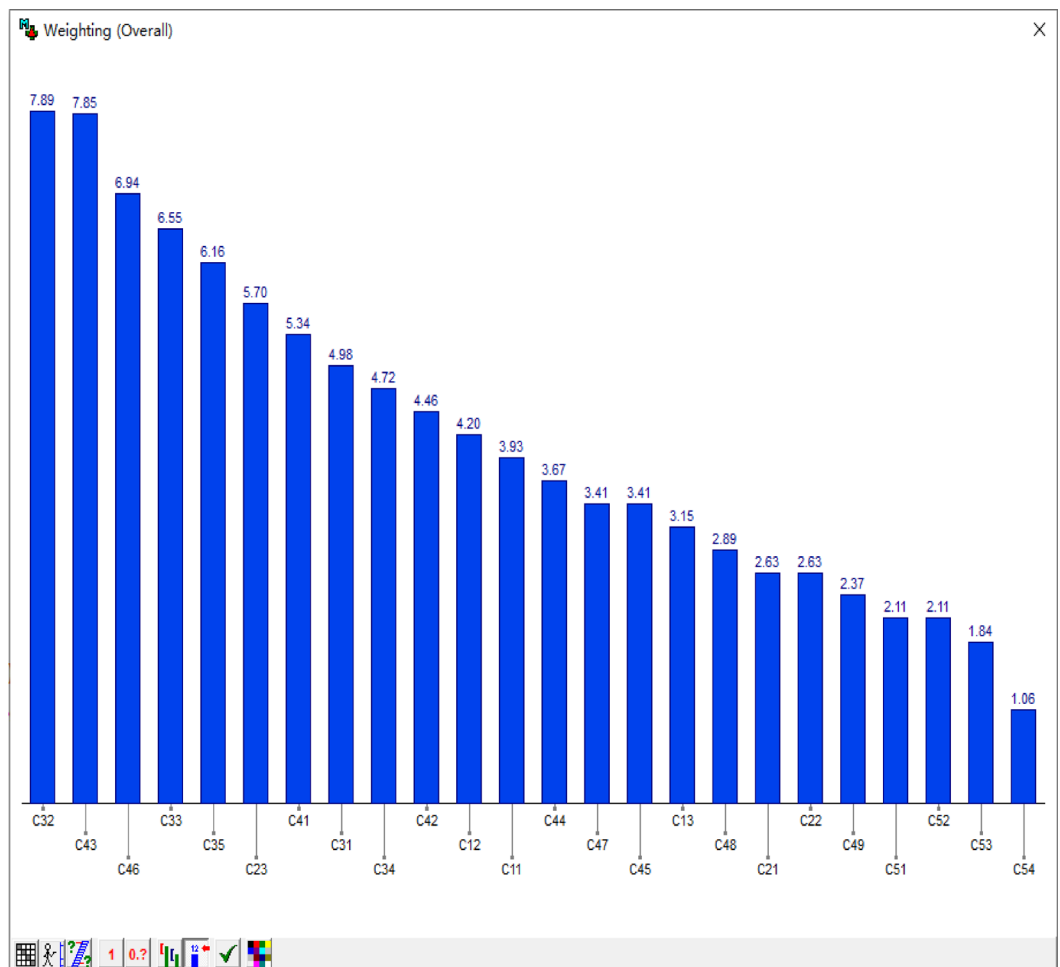


Fig. 4. The generated criteria weights for teaching faculty.

By following the aforementioned steps for the other two faculty career tracks, the criteria weights for the studied college can be derived, as shown in Table 5. Correspondingly, the top three most significant criteria in performance evaluation for research faculty are publication records and patent (C_{51}), scientific research projects (C_{52}), and scientific research awards (C_{53}). In terms of teaching-research balanced faculty, the key evaluation indicators with higher weights include first-class curriculum development (C_{43}), publication records and patent (C_{51}), teaching achievement awards (C_{41}), and scientific research awards (C_{53}).

Ranking the evaluation results

In College X, the director of the Teaching and Research Office is responsible for collecting data related to the performance criteria of all faculty members and verifying the accuracy of the results. Subsequently, all faculty members are ranked in a single pool based on their total scores to determine the recipients of outstanding faculty awards. The disadvantage of the current method is that it fails to differentiate the importance of various criteria and lacks a basis for an apples-to-apples comparison in ranking.

For illustrative purposes, we take 20 teaching faculty members from the college as an example to demonstrate the proposed ranking procedure. The evaluated faculty members are denoted as F_1, F_2, \dots, F_{20} . Their performance data for 2024 were provided by the director of the Teaching and Research Office. Regarding missing data resulting from the additional criteria in the proposed framework, we requested the decision-making group to conduct an evaluation to ensure data completeness.

Based on the collected data, the boundaries were calculated in accordance with Step 2 in Subsection 3.2, and the results are presented in Table 6. It was noted that none of the faculty members had achievements in three areas in 2024, including outstanding graduation thesis awards (C_{48}), laboratory construction (C_{49}), and scientific research projects (C_{52}). Consequently, the boundaries were not applicable to these three criteria and 0 was used as the result for each. Next, by following Steps 3 to 5 in Subsection 3.2, a restricted transition decision matrix m_{ij} was created, as depicted in Figure A2 of the Supplementary Figures. Using the weights obtained from the MACBETH approach, a weighted and corrected decision matrix s_{ij} was generated, along with the final ranking, in accordance with Step 6 in Fuzzy filter ranking. These are presented in Figure A3 of the Supplementary Figures. Ultimately, faculty members F_{11} , F_{16} , and F_{19} identified as the top three performers in the teaching faculty track, as shown in Fig. 5.

Metrics	Teaching faculty	Research faculty	Teaching-research balanced faculty
Political and ideological performance (C11)	3.93	4.05	4.07
Professional ethics and literacy (C12)	4.20	4.46	4.21
Teamwork awareness (C13)	3.15	3.47	4.13
Service work (C21)	2.63	3.96	2.94
Coach junior lecturer (C22)	2.63	3.68	2.77
Industry engagement (C23)	5.70	1.32	3.67
Teaching workload (C31)	4.98	2.87	3.35
Student evaluation (C32)	7.89	3.56	5.75
Peer evaluation (C33)	6.55	3.56	3.54
International student class (C34)	4.72	0.19	1.80
Teaching rules (C35)	6.16	5.18	5.96
Teaching achievement awards (C41)	5.34	2.65	6.24
Teaching reform projects (C42)	4.46	2.65	5.43
First-class curriculum development (C43)	7.85	2.76	6.93
Publication of textbooks and articles in teaching (C44)	3.67	5.90	4.87
Mentoring provincial college students' innovation program (C45)	3.41	6.14	1.24
Teaching skill competitions (C46)	6.94	1.44	5.16
Instruct students in various competitions (C47)	3.41	2.03	3.86
Outstanding graduation thesis award (C48)	2.89	6.30	2.15
Laboratory construction (C49)	2.37	0.73	0.75
Publication records and patent (C51)	2.11	9.35	6.87
Scientific research projects (C52)	2.11	8.93	5.43
Scientific research awards (C53)	1.84	8.83	6.24
Horizontal scientific research projects (C54)	1.06	5.99	2.64

Table 5. The obtained criteria weights.

	A_i^{max}	A_i^{min}	A_i^{avg}	A_i^{Uh}	A_i^{Ul}	A_i^{Lh}	A_i^{Ll}
C11	3	2	2.1	2.64	2.28	2.08	2.04
C12	3	1	2.05	2.62	2.24	1.84	1.42
C13	4	1	2.15	3.26	2.52	1.92	1.46
C21	3	1	2	2.6	2.2	1.8	1.4
C22	3	2	2.15	2.66	2.32	2.12	2.06
C23	5	1	2.25	3.9	2.8	2	1.5
C31	507	205	352	445	383	322.6	263.8
C32	95.30833	89.785	93.00339	94.38635	93.46438	92.35971	91.07235
C33	10	9	9.85	9.94	9.88	9.68	9.34
C34	3	0	0.35	1.94	0.88	0.28	0.14
C35	2	0	0.2	1.28	0.56	0.16	0.08
C41	100	0	8	63.2	26.4	6.4	3.2
C42	15	0	1.5	9.6	4.2	1.2	0.6
C43	20	0	1.15	12.46	4.92	0.92	0.46
C44	2	0	0.2	1.28	0.56	0.16	0.08
C45	6	0	0.5	3.8	1.6	0.4	0.2
C46	8	0	0.7	5.08	2.16	0.56	0.28
C47	52	0	6.3	33.72	15.44	5.04	2.52
C48	0	0	0	0	0	0	0
C49	0	0	0	0	0	0	0
C51	5	0	0.35	3.14	1.28	0.28	0.14
C52	0	0	0	0	0	0	0
C53	7	0	0.35	4.34	1.68	0.28	0.14
C54	30,000	0	2500	19,000	8000	2000	1000

Table 6. boundaries.

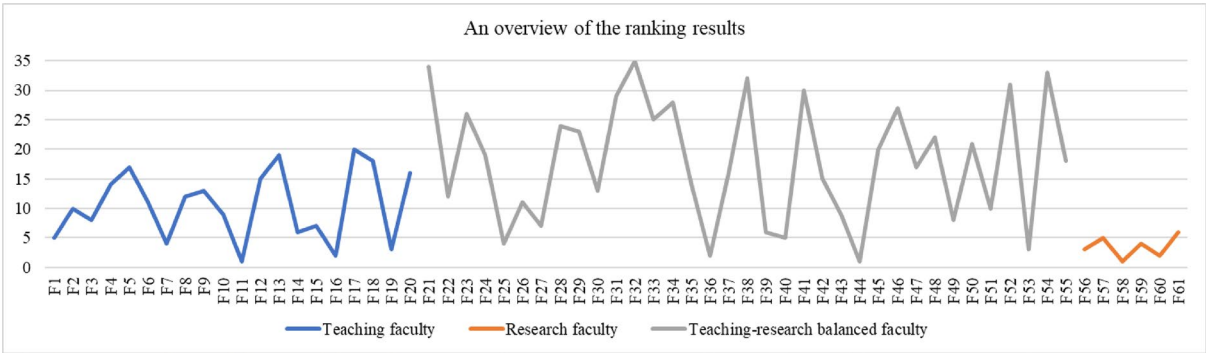


Fig. 5. An overview of the ranking results.

Main Criteria	Original weights	Modified weights of all criteria when weight of teaching achievements is varied from 0.1 to 0.9								
C1	0.1128	0.1702	0.1513	0.1323	0.1134	0.0945	0.0756	0.0567	0.0378	0.0189
C2	0.1096	0.1653	0.1470	0.1286	0.1102	0.0919	0.0735	0.0551	0.0367	0.0184
C3	0.3030	0.4571	0.4063	0.3555	0.3047	0.2539	0.2032	0.1524	0.1016	0.0508
C4	0.4034	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
C5	0.0712	0.1074	0.0955	0.0835	0.0716	0.0597	0.0477	0.0358	0.0239	0.0119

Table 7. Weights of main criteria in sensitivity analysis.

Similarly, teaching-research balanced faculty are denoted as $F_{21}, F_{22} \dots F_{55}$, and research faculty as $F_{56}, F_{57} \dots F_{61}$. By repeating the process elaborated above, the ranking results for these two faculty tracks can be obtained, as shown in Fig. 5. This ranking procedure is highly user-friendly, as it merely requires the use of pre-defined formulas, eliminating the need for manual calculations. Based on the faculty population in each track, outstanding faculty awards can be granted to the top 3 teaching faculty, top 1 research faculty, and top 6 teaching-research balanced faculty. These results align with the dean’s perceptions, thereby validating the effectiveness of the proposed model.

Sensitivity analysis

To verify the robustness of the solution, a sensitivity analysis is conducted to examine how changes in the weights of evaluation criteria affect the performance evaluation results for teaching faculty. Among all main criteria, teaching achievements (C_4) have the highest relative weight, so they are likely to significantly influence the global weights of the performance criteria. For the sensitivity analysis, following the suggestion by Talib et al.⁴⁹, the weight of the most important criterion (teaching achievements) is varied from 0.1 to 0.9 in increments of 0.1, with a proportionate adjustment made to the weights of all other criteria. The changes in the weights of the main criteria during the sensitivity analysis are presented in Table 7.

The changes in the weights of the main criteria result in adjustments to the global weights and rankings of the performance criteria, presented in Tables 8 and 9, respectively. As evident from Tables 8 and 9, when the weight of the main criterion “teaching achievements” ranges from 0.1 to 0.4, the sub-criterion “student evaluation” (C_{32}) ranks first. In Scenarios 1–3, where the main criterion “teaching activities” (C_3) holds the highest weight, its sub-criteria C_{32} , C_{33} , and C_{35} are among the top three. In Scenario 4, the top three performance criteria are C_{32} , C_{43} , and C_{46} . However, when the weight of “teaching achievements” (C_4) ranges from 0.5 to 0.9, the performance criteria C_{43} , C_{46} , and C_{41} rank 1st, 2nd, and 3rd. This shift in the order of key indicators is logical, as their parent criterion “teaching achievements” (C_4) carries the highest relative weight in these scenarios.

Changes in the global weights of the criteria also lead to changes in the performance rankings of teaching faculty, as shown in Table 10. As observed, the top-ranked faculty member (F_{11}) and the bottom three (F_{18} , F_{13} , and F_{17}) remain unchanged despite weight adjustments. Generally, these faculty members are regarded as the top performer and the weakest performers, respectively, with weight changes having no impact on their rankings. In Scenarios 1–4, faculty member F_{16} ranks 2nd but drops to 3rd to 7th in Scenarios 5–9. This result is reasonable because this faculty member performs stronger in the criterion “teaching activities” (C_3) yet weaker in the criterion “teaching achievements” (C_4). Consequently, her ranking is higher in the first four scenarios, where the main criterion C_3 carries the highest weight.

Discussion

The results of this case study, which applies the integrated MACBETH-FFR framework, demonstrate its effectiveness in evaluating faculty performance at application-oriented universities. The MACBETH method, with its capacity to process qualitative judgments and account for the relative importance of criteria, provides a more nuanced approach to weight determination. This is particularly crucial given that faculty members in

Sub-criteria	Original weights	Global weights of indicators when the weight of teaching achievement is varied from 0.1 to 0.9 in different scenarios (S1–S9)								
		S1 (0.1)	S2 (0.2)	S3 (0.3)	S4 (0.4)	S5 (0.5)	S6 (0.6)	S7 (0.7)	S8 (0.8)	S9 (0.9)
C11	0.0393	0.0593	0.0527	0.0461	0.0395	0.0329	0.0263	0.0198	0.0132	0.0066
C12	0.042	0.0634	0.0563	0.0493	0.0422	0.0352	0.0282	0.0211	0.0141	0.0070
C13	0.0315	0.0475	0.0422	0.0370	0.0317	0.0264	0.0211	0.0158	0.0106	0.0053
C21	0.0263	0.0397	0.0353	0.0309	0.0264	0.0220	0.0176	0.0132	0.0088	0.0044
C22	0.0263	0.0397	0.0353	0.0309	0.0264	0.0220	0.0176	0.0132	0.0088	0.0044
C23	0.057	0.0860	0.0764	0.0669	0.0573	0.0478	0.0382	0.0287	0.0191	0.0096
C31	0.0498	0.0751	0.0668	0.0584	0.0501	0.0417	0.0334	0.0250	0.0167	0.0083
C32	0.0789	0.1190	0.1058	0.0926	0.0793	0.0661	0.0529	0.0397	0.0264	0.0132
C33	0.0655	0.0988	0.0878	0.0769	0.0659	0.0549	0.0439	0.0329	0.0220	0.0110
C34	0.0472	0.0712	0.0633	0.0554	0.0475	0.0396	0.0316	0.0237	0.0158	0.0079
C35	0.0616	0.0929	0.0826	0.0723	0.0620	0.0516	0.0413	0.0310	0.0207	0.0103
C41	0.0534	0.0132	0.0265	0.0397	0.0529	0.0662	0.0794	0.0927	0.1059	0.1191
C42	0.0446	0.0111	0.0221	0.0332	0.0442	0.0553	0.0663	0.0774	0.0884	0.0995
C43	0.0785	0.0195	0.0389	0.0584	0.0778	0.0973	0.1168	0.1362	0.1557	0.1751
C44	0.0367	0.0091	0.0182	0.0273	0.0364	0.0455	0.0546	0.0637	0.0728	0.0819
C45	0.0341	0.0085	0.0169	0.0254	0.0338	0.0423	0.0507	0.0592	0.0676	0.0761
C46	0.0694	0.0172	0.0344	0.0516	0.0688	0.0860	0.1032	0.1204	0.1376	0.1548
C47	0.0341	0.0085	0.0169	0.0254	0.0338	0.0423	0.0507	0.0592	0.0676	0.0761
C48	0.0289	0.0072	0.0143	0.0215	0.0287	0.0358	0.0430	0.0501	0.0573	0.0645
C49	0.0237	0.0059	0.0118	0.0176	0.0235	0.0294	0.0353	0.0411	0.0470	0.0529
C51	0.0211	0.0318	0.0283	0.0248	0.0212	0.0177	0.0141	0.0106	0.0071	0.0035
C52	0.0211	0.0318	0.0283	0.0248	0.0212	0.0177	0.0141	0.0106	0.0071	0.0035
C53	0.0184	0.0278	0.0247	0.0216	0.0185	0.0154	0.0123	0.0093	0.0062	0.0031
C54	0.0106	0.0160	0.0142	0.0124	0.0107	0.0089	0.0071	0.0053	0.0036	0.0018

Table 8. Weights of performance criteria in sensitivity analysis.

Sub-criteria	Original rank	S1 (0.1)	S2 (0.2)	S3 (0.3)	S4 (0.4)	S5 (0.5)	S6 (0.6)	S7 (0.7)	S8 (0.8)	S9 (0.9)
C11	12	8	8	10	12	16	17	17	17	17
C12	11	7	7	9	11	15	16	16	16	16
C13	16	9	9	12	16	18	18	18	18	18
C21	18	10	11	14	18	19	19	19	19	19
C22	18	10	11	14	18	19	19	19	19	19
C23	6	4	4	4	6	8	12	13	13	13
C31	8	5	5	5	8	12	14	14	14	14
C32	1	1	1	1	1	4	6	10	10	10
C33	4	2	2	2	4	6	9	11	11	11
C34	9	6	6	7	9	13	15	15	15	15
C35	5	3	3	3	5	7	11	12	12	12
C41	7	18	16	11	7	3	3	3	3	3
C42	10	19	18	13	10	5	4	4	4	4
C43	2	15	10	6	2	1	1	1	1	1
C44	13	20	19	16	13	9	5	5	5	5
C45	14	21	20	17	14	10	7	6	6	6
C46	3	16	13	8	3	2	2	2	2	2
C47	14	21	20	17	14	10	7	6	6	6
C48	17	23	22	22	17	14	10	8	8	8
C49	20	24	24	23	20	17	13	9	9	9
C51	21	12	14	19	21	21	21	21	21	21
C52	21	12	14	19	21	21	21	21	21	21
C53	23	14	17	21	23	23	23	23	23	23
C54	24	17	23	24	24	24	24	24	24	24

Table 9. The rank of performance criteria for teaching faculty in sensitivity analysis.

Faculty	Original rank	S1 (0.1)	S2 (0.2)	S3 (0.3)	S4 (0.4)	S5 (0.5)	S6 (0.6)	S7 (0.7)	S8 (0.8)	S9 (0.9)
F1	5	4	4	5	5	5	6	5	5	5
F2	10	9	10	9	10	10	10	10	9	9
F3	8	5	5	7	8	9	9	9	10	10
F4	14	15	15	15	14	13	13	13	13	13
F5	17	16	17	17	17	17	17	17	17	17
F6	11	11	12	11	11	11	11	11	11	11
F7	4	3	3	3	4	6	7	8	8	8
F8	12	12	13	14	12	12	12	12	12	12
F9	13	6	9	12	13	14	14	14	14	14
F10	9	17	14	10	9	8	8	7	6	6
F11	1	1	1	1	1	1	1	1	1	1
F12	15	7	11	13	15	15	15	15	15	15
F13	19	19	19	19	19	19	19	19	19	19
F14	6	13	8	8	6	4	3	3	3	3
F15	7	10	7	6	7	7	5	4	4	4
F16	2	2	2	2	2	3	4	6	7	7
F17	20	20	20	20	20	20	20	20	20	20
F18	18	18	18	18	18	18	18	18	18	18
F19	3	8	6	4	3	2	2	2	2	2
F20	16	14	16	16	16	16	16	16	16	16

Table 10. The rank of teaching faculty in sensitivity analysis.

different career tracks have distinct priorities. For instance, teaching faculty may place more emphasis on teaching-related criteria, while research faculty focus more on research outputs, which is reflected in the relatively higher weights assigned to the corresponding criteria. The FFR method, on the other hand, offers a simple yet reliable means of ranking faculty performance. Its use of discrete fuzzy scores and filters streamlines the ranking process, making it accessible even to individuals without a strong mathematical background. A significant advantage is its ability to accommodate diverse measurement units across criteria and bypass verification of the consistency of estimates. Higher education institutions can also set minimum achievement thresholds for specific criteria and adjust boundaries as needed.

Moreover, analyzing the results helps the university identify areas for improvement in faculty management, such as enhancing faculty training in research or encouraging more active participation in teaching skill competitions. If a large number of faculty members fail to meet the first filter—for example, first-class curriculum development (C_{43})—it indicates that the college may need to intensify efforts in this area by organizing experience-sharing sessions and knowledge transfer workshops. This not only helps faculty improve their teaching skills but also enriches teaching content, better preparing students for their professional careers.

From the decision-making group’s perspective, the integrated model provides a robust and adaptable assessment framework for faculty performance evaluation, enabling the school to navigate complexities strategically and effectively. The user-friendliness of the tools and the streamlined process enhance the practicality of this framework. The decision-making group also expressed interest in implementing the proposed methodology in the faculty performance evaluation to be conducted next year.

Concluding remarks

Faculty performance evaluation in higher education institutions is critical for ensuring teaching quality, enhancing student learning outcomes, and fostering continuous professional development. Reforming and innovating faculty performance evaluation systems in application-oriented universities is vital to the sustainable development of such institutions. This paper proposes an integrated multi-criteria evaluation framework, providing an alternative methodology for evaluating faculty performance in application-oriented universities. The proposed assessment methodology combines two components: the MACBETH method, which determines a weighting scheme based on semantic scales, and the FFR approach, which uses the weights generated by MACBETH to rank faculty members. By considering the unique characteristics of application-oriented universities and differentiating evaluation focuses according to faculty career tracks, the framework addresses the limitations of traditional evaluation methods.

The case study validates the effectiveness of the proposed framework. It provides a more scientific, comprehensive, and objective evaluation of faculty performance, enabling universities to make more informed decisions in aspects of faculty management, such as recruitment, retention, and professional development. This study contributes to the field of faculty performance evaluation in application-oriented universities by providing a practical and effective evaluation tool. Notably, the framework can be extended to research-oriented universities, thereby facilitating the high-quality development of higher education. Additionally, the methodology can be adapted to address problems such as university ranking and academic staff selection. The

combination of MACBETH for weight determination and FFR for ranking offers a user-friendly and reliable solution that is easy to implement.

In terms of future research, given that faculty evaluation systems may vary across countries, further studies can be conducted to refine and supplement the initial set of indicators by incorporating insights and perspectives from a broader pool of experts. Additionally, future research could develop a mini-program that can either be embedded into a module of the university's administrative software or operate independently to facilitate the performance evaluation process. Furthermore, researchers may explore integrating this framework with other emerging evaluation methods or technologies to enhance its accuracy and adaptability. For instance, big data analytics could be incorporated to include more diverse data sources in the evaluation process.

Data availability

The data that supports the findings of this study are partially included in this article. The complete data are available from the corresponding author upon request.

Received: 23 April 2025; Accepted: 25 August 2025

Published online: 27 August 2025

References

1. Khurram, R. & Ata, M. Evaluating the effectiveness of performance appraisal systems in higher education: A Faculty-Centric study. *Propel J. Appl. Manag.* **4** (1), 1–25 (2024).
2. Cadez, S., Dimovski, V. & Zaman Groff, M. Research, teaching and performance evaluation in academia: the salience of quality. *Stud. High. Educ.* **42** (8), 1455–1473 (2017).
3. Besancenot, D., Faria, J. R. & Vranceanu, R. Why business schools do so much research: A signaling explanation. *Res. Policy.* **38** (7), 1093–1101 (2009).
4. Ter Bogt, H. J. & Scapens, R. W. Performance management in universities: effects of the transition to more quantitative measurement systems. *Eur. Acc. Rev.* **21** (3), 451–497 (2012).
5. Goodwin, P. & Wright, G. *Decision Analysis for Management Judgment* (Wiley, 2014).
6. Carnero, M. C. Developing a fuzzy TOPSIS model combining MACBETH and fuzzy Shannon entropy to select a gamification app. *Mathematics* **9** (9), 1034 (2021).
7. Chang, T. Y. & Ku, C. C. Y. Fuzzy filtering ranking method for multi-criteria decision making. *Comput. Ind. Eng.* **156**, 107217 (2021).
8. Wang, X., Ferreira, F. A., Tao, M. & Chang, C. T. A hybrid AHP–FCE–WMCGP approach for internal auditor selection: A generic framework. *Int. J. Fuzzy Syst.* **24** (5), 2229–2249 (2022).
9. Bernroider, E. W. N. & Schmöller, P. A technological, organisational, and environmental analysis of decision making methodologies and satisfaction in the context of IT induced business transformations. *Eur. J. Oper. Res.* **224** (1), 141–153 (2013).
10. Štilić, A. & Puška, A. Integrating Multi-Criteria Decision-Making methods with sustainable engineering: A comprehensive review of current practices. *Eng* **4** (2), 1536–1549 (2023).
11. Badiee, A., Moshtari, M. & Berenguer, G. A systematic review of operations research and management science modeling techniques in the study of higher education institutions. *Socio-Econ. Plann. Sci.* **93**, 101889 (2024).
12. Yüksel, F. Ş., Kayadelen, A. N. & Antmen, F. A systematic literature review on multi-criteria decision making in higher education. *Int. J. Assess. Tools Educ.* **10** (1), 12–28 (2023).
13. Afriadi, B. Multiple criteria decision making (MCDM) for quality education management in higher education: systematic review. *Int. J. Bus. Law Educ.* **3** (2), 173–180 (2022).
14. Johnes, J. Operational research in education. *Eur. J. Oper. Res.* **243** (3), 683–696 (2015).
15. Malik, D. A. A., Yusof, Y., Na'im, K. & Khalif, K. M. A view of MCDM application in education. In *J. Phys. Conf. Series* (Vol. 1988, No. 1, p. 012063). IOP Publishing. (2021), July.
16. Smilowitz, K. & Keppler, S. On the use of operations research and management in public education systems. *Pushing Boundaries: Front. Impactful OR/OM Research*, 84–105. (2020).
17. Sinuany-Stern, Z. Operations research and management science in higher education: an overview. *Handbook of Operations Research and Management Science in Higher Education*, 3–34. (2021).
18. Badri, M. A. & Abdulla, M. H. Awards of excellence in institutions of higher education: an AHP approach. *Int. J. Educational Manage.* **18** (4), 224–242 (2004).
19. Feng, Y. J., Lu, H. & Bi, K. An AHP/DEA method for measurement of the efficiency of R&D management activities in universities. *Int. Trans. Oper. Res.* **11** (2), 181–191 (2004).
20. Bougnol, M. L. & Dulá, J. H. Validating DEA as a ranking tool: an application of DEA to assess performance in higher education. *Ann. Oper. Res.* **145**, 339–365 (2006).
21. Nikoomaram, H., Mohammadi, M., Taghipourian, M. J. & Taghipourian, Y. Training performance evaluation of administration sciences instructors by fuzzy MCDM approach. *Contemp. Eng. Sci.* **2** (12), 559–575 (2009).
22. Wu, H. Y., Chen, J. K., Chen, I. S. & Zhuo, H. H. Ranking universities based on performance evaluation by a hybrid MCDM model. *Measurement* **45** (5), 856–880 (2012).
23. Kiarazm, A. & Koohkan, F. Performance evaluation in higher education institutes with the use of combinative model AHP and BSC. *J. Basic Appl. Sci. Res.* **3** (4), 940–944 (2013).
24. Aly, M. F., Attia, H. A. & Mohammed, A. M. Prioritizing faculty of engineering education performance by using AHP-TOPSIS and balanced scorecard approach. *Int. J. Eng. Sci. Innovative Technol.* **3** (1), 11–23 (2014).
25. Thanassoulis, E., Dey, P. K., Petridis, K., Goniadis, I. & Georgiou, A. C. Evaluating higher education teaching performance using combined analytic hierarchy process and data envelopment analysis. *J. Oper. Res. Soc.* **68** (4), 431–445 (2017).
26. El Gibari, S., Gómez, T. & Ruiz, F. Evaluating university performance using reference point based composite indicators. *J. Informetrics* **12** (4), 1235–1250 (2018).
27. Blasco-Blasco, O., Liern-García, M., López-García, A. & Parada-Rico, S. E. An academic performance indicator using flexible multi-criteria methods. *Mathematics* **9** (19), 2396 (2021).
28. Ma, Z., See, K. F., Yu, M. M. & Zhao, C. Research efficiency analysis of china's university faculty members: A modified meta-frontier DEA approach. *Socio-Economic Plann. Sci.* **76**, 100944 (2021).
29. Villegas, J. G., Castañeda P. C. & Castañeda Gómez, E. Planning and performance measurement in higher education: three case studies of operational research application. *Revista Facultad De Ingeniería Universidad De Antioquia*. **100**, 97–112 (2021).
30. Petrudi, S. H. H., Ghomi, H. & Mazaheriasad, M. An integrated fuzzy Delphi and best worst method (BWM) for performance measurement in higher education. *Decis. Analytics J.* **4**, 100121 (2022).
31. Ayyildiz, E., Murat, M., Imamoglu, G. & Kose, Y. A novel hybrid MCDM approach to evaluate universities based on student perspective. *Scientometrics* **128** (1), 55–86 (2023).

32. Karahan, M. & Karahan, M. S. Comparative analysis of the academic performance of research universities in Türkiye using MCDM techniques. *Yükseköğretim Dergisi*. **15** (1), 195–208 (2025).
33. Do, Q. H. Evaluating lecturer performance in vietnam: an application of fuzzy AHP and fuzzy TOPSIS methods. *Heliyon* **10**, e30772 (2024).
34. Karmaker, C. L. et al. A framework of faculty performance evaluation: A case study in Bangladesh. *Int. J. Res. Adv. Eng. Technol.* **4** (3), 18–24 (2018).
35. Hafizah, H., Napitupulu, D., Adiyarta, K. & Windarto, A. P. Decision support system for teacher performance assessment of SMK Nusantara 1 Ciputat based on AHP and TOPSIS. In *Journal of Physics: Conference Series* (Vol. 1255, No. 1, p. 012055). IOP Publishing, (2019), August.
36. Do, A. D. et al. A dynamic fuzzy multiple criteria decision-making approach for lecturer performance evaluation. *J. Manage. Inform. Decis. Sci.* **22** (3), 250–261 (2019).
37. Daniawan, B. Evaluation of lecturer teaching performance using AHP and SAW methods. *bit-Tech* **1** (2), 74–83 (2018).
38. Radovanović, M., Jovčić, S., Petrovski, A. & Cirkin, E. Evaluation of university professors using the spherical fuzzy AHP and grey MARCOS Multi-Criteria Decision-Making model: A case study. *Spectr. Decis. Mak. Appl.* **2** (1), 198–218 (2025).
39. Aazagreyir, P. et al. B. I. S. M. A. R. K. Fuzzy DEM-SAW: A novel hybridized model of fuzzy DEMATEL-SAW in lecturers' performance evaluation based on teaching, research, service and commercialization (TRSC) criteria for promotion. *Journal Theoretical Appl. Inform. Technology*, **102**(11). (2024).
40. Watrianthos, R., Ritonga, W. A., Rengganis, A., Wanto, A. & Indrawan, M. I. Implementation of PROMETHEE-GAIA Method for Lecturer Performance Evaluation. In *Journal of Physics: Conference Series* (Vol. 1933, No. 1, p. 012067). IOP Publishing, (2021), June.
41. Agrell, P. J. & Steuer, R. E. ACADDEA—A decision support system for faculty performance reviews. *J. Multi-Criteria Decis. Anal.* **9** (5), 191–204 (2000).
42. Qian, J., Zhou, G., He, W., Cui, Y. & Deng, H. Optimization of teacher evaluation indicator system based on fuzzy-DEMATEL-BP. *Heliyon*, **10**(13). (2024).
43. e Costa, C. A. B. & Vansnick, J. C. MACBETH—An interactive path towards the construction of Cardinal value functions. *Int. Trans. Oper. Res.* **1** (4), 489–500 (1994).
44. Ferreira, F. A. F. & Santos, S. P. Two decades on the MACBETH approach: A bibliometric analysis. *Ann. Oper. Res.* **296**, 901–925 (2021).
45. e Costa, C. A. B. & Oliveira, M. D. A multicriteria decision analysis model for faculty evaluation. *Omega* **40** (4), 424–436 (2012).
46. Purwani, F. An analytic network process method approach to design models of lecturers performance evaluation. *Inter. J. Artif. Intell. Res.*, **6**(1.2). (2024).
47. Retnowati, T. H., Mardapi, D., Kartowagiran, B. & Hamdi, S. A model of lecturer performance evaluation: sustainable lecturer performance mapping. *Int. J. Instruction.* **14** (2), 83–102 (2021).
48. Buyukdagli, O. & Yeralan, S. How do weights affect faculty performance evaluations? *Sustain. Eng. Innov.* **2** (2), 89–101 (2020).
49. Talib, F., Asjad, M., Attri, R., Siddiquee, A. N. & Khan, Z. A. Ranking model of total quality management enablers in healthcare establishments using the best-worst method. *TQM J.* **31** (5), 790–814 (2019).

Author contributions

M.T. and Q.C. led the analysis and writing of the paper, W.Y. prepared figures and operated software, and X.W. led the project and oversaw the project progress. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Compliance with ethical standards

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-17537-6>.

Correspondence and requests for materials should be addressed to X.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025