Heliyon 9 (2023) e13134

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

A combination of FDM, EWM, and GRA to measure performance in higher vocational college: An empirical case from China

Lei-Yi Peng [a,*], Le-Ying Pan [b], Jia Lu [a], Yan-Man Long [c]

[a] *Faculty of Tourism, Zhejiang Business College, Hangzhou, 310053, Zhejiang, China*
[b] *Department of Foreign Language, Zhejiang Business College, Hangzhou, 310053, Zhejiang, China*
[c] *Students Mental Health Counselling Centre, Kaili University, Kaili, 555600, Guizhou, China*

A R T I C L E I N F O

Keywords:
Higher vocational college
Performance measurement
FDM
EWM
GRA

A B S T R A C T

Given the problems of unclear division of evaluation factors, inadequate utilization of objective data and unreasonable distribution of weight in college performance measurement, this study extracts relevant indicators and quantitative data from the survey report of administrational institutions, and constructs a comprehensive performance measurement model based on the integration of fuzzy Delphi method (FDM), Entropy weight method (EWM), and grey relational analysis (GRA). The study seeks to identify differences in measurement analysis by comparing the weights, the performance and the rankings of the alternatives about higher vocational colleges in Zhejiang province. Finally, the results show the following: (1) by FDM, 33 indicators about colleges' performance are selected to form the evaluation system of the college's performance, (2) the indicators' weights are obtained through EWM and GRA, and a nonparametric test shows that there is no significant difference between the two types of weights, (3) the grey relational degrees of the alternatives are obtained and ranked on the basis of comprehensive evaluation model. By nonparametric test, there is a significant difference between the two types of relational degrees. On the contrary, no significant difference is found in the ranking of relational degrees, (4) based on the analysis results, this study further compares the performance of alternatives in different forms. There are significant differences between the performance of public colleges and private colleges, while no significant differences are observed between the performance of vocational colleges in Hangzhou and non-Hangzhou. Given the reliability and validity of the model, the comprehensive measurement model provides a relatively objective reference for college governance and administrational institutions, and also becomes an effective tool of colleges' evaluation to assist and improve the management practice in educational institutions.

1. Introduction

Organizational performance is regarded as an essential variable to measure organizational success [1]. The concept of organization is also considered synonymous with business, so that educational institutions can be categorized into one type of organization [2]. Although, the higher education has a significant expansion about the number of students and academic affiliates in countries over recent decades, this expansion has not been matched with an improvement in the quality of education [3]. Performance evaluation is a

* Corresponding author.
E-mail addresses: 00373@zjbc.edu.cn (L.-Y. Peng), 00124@zjbc.edu.cn (L.-Y. Pan), 00461@zjbc.edu.cn (J. Lu), longyanman@kluniv.edu.cn (Y.-M. Long).

https://doi.org/10.1016/j.heliyon.2023.e13134
Received 26 July 2022; Received in revised form 15 January 2023; Accepted 18 January 2023
Available online 21 January 2023

fundamental and important process for all organizations, especially for the higher education system, and its goal is to achieve excellence and improve quality in teaching-learning processes [4]. However, academia and industry have never reached a complete consensus on the definition of organizational achievement and evaluation. As one of the applications of performance measurement, the ranking of universities has been in great demand, but there are also great controversies [5], because the performance measurement of universities triggered plenty of doubts about the evaluation criteria, evaluation data, evaluation methods, and application of evaluation results. Therefore, the attention of many researchers has been attracted to the field of performance evaluation of universities and higher education institutions because of the critical role of these institutions in the society and the existence of a competitive market among universities [6].

As higher education in China has entered the stage of popularization in a short period, higher vocational education plays an indispensable role. Simultaneously, Zhejiang province is one of the regions where higher vocational education flourishes, and higher vocational colleges account for half of the number of the province's colleges and universities. In 2014–2015, there were 46 independent higher vocational colleges in the province, wherein there were 37 public colleges and 9 private colleges with 343,200 full-time students. The total fixed assets of higher vocational colleges totaled CNY 25.3 billion, and funds for running all colleges reached CNY 8.801 billion. Compared with other provinces, the student-teacher ratio jumped to 11th, the per-student financial special investment rose to 7th, the annual per-student financial allocation level ranked 5th, and the per-student teaching equipment listed 6th. Thus, the overall education funding guarantee index has basically remained in the country's top tier. Moreover, the proportion of double-qualified teachers ranked 1st, which has obvious advantages over other provinces. The number of full-time teachers per college took 8th spot. Through the study on the performance of higher vocational colleges in Zhejiang, we can find the benchmark for the development as well as explore the shortcomings that affect the development of higher vocational education. More importantly, the paper aims to provide an effective tool about colleges' evaluation to assist and enhance the management practice in educational institutions under the background of "*Shuang Gao*", which is a major construction project about higher vocational colleges and key majors (groups). The initiation of the project is to build up a vocational education system with Chinese characteristics and world-class standards, as well as to promote the high-level development of vocational education in China context.

Universities need to choose an appropriate method for evaluating performance precisely to improve the level of accountability and customer orientation [7]. Common methods of performance evaluation include multivariate analysis, data employment analysis (DEA), and multi-criteria decision making (MCDM), among which there are both single methods and hybrid methods. No method can be considered as a best method either for a general or for a particular problem [8], while the hybrid MCDM method can achieve a certain level of optimization through the application of multiple MCDM. Integrating MCDM with fuzzy theory or grey theory is the preferred way to make decision models close to real-life problems, because fuzzy logic could help to overcome uncertainties that arise from human qualitative judgments and incomplete preference relationships [9]. Furthermore, the combination of multiple evaluation methods has become another hot spot in today's comprehensive evaluation methods [10]. Additionally, a system's performance is analyzed, quantitative performance measurement are preferred to the qualitative assessments because of the uncertainty and difficulties [11]. Hence, given that the widespread use of evaluation methods suggests that the key to university classification and ranking is model selection [5], constructing a comprehensive evaluation model based on objective weight methods has become another motivation for this study. This paper postulates a hybrid approach based on fuzzy Delphi method (FDM), Entropy weight method (EWM) and grey relation analysis (GRA) to measure the performance of higher vocational colleges. More specifically, this study answers the following research questions (RQ):

RQ1. What are the main criteria to measure performance of higher vocational colleges?

RQ2. What are the criteria' weights of the two methods, and what's the difference of weights?

RQ3. What are the relational degrees of the alternatives, and what's the difference of relational degrees and ranking of relational degrees?

The next section reviews literature about performance evaluation of higher education, and MCDM techniques. Section 3 proposes a comprehensive evaluation model. And then, section 4 illustrates the results of empirical analysis, and section 5 conducts the discussion. Finally, section 6 briefly explains the concluding statement.

2. Literature review

2.1. Evaluation of higher education

Evaluation is applied in all levels of education within the field of higher education. It is implemented in many aspects such as: evaluation of universities, evaluation of study programs, evaluation of academic staff, and teaching evaluation. Different countries place their own emphasis, and many terms are frequently used as synonyms for measurement, including achievement control, effect control, and quality control, efficiency research and assessment research, etc. This implies not only that the scholars mainly vary in the emphasis of their study topics, but also that there is still a lack of consistent consensus on a single definition of performance measurement [12].

From the angle of measurement methods, the studies mainly focus on feedback, formative evaluation, and peer evaluation [13]. Wu et al. (2012) proposed a combined approach using Analytic Hierarchy Process (AHP) and VlseKriterijumskaOptimizacija I KompromisnoResenje (VIKOR) to rank 12 private universities in Taiwan [14]. Zolfani and Ghadikolaei (2013) applied DEMATEL to clarify cause and effect relations of perspectives of balanced scorecard, and Analytic Network Process (ANP) to calculate weights of indicators in perspectives and VIKOR to rank five universities in Iran [15]. Bayraktar et al. (2013) combined DEA and Stochastic Frontier Analysis (SFA) to evaluate the quality control efficiency of relevant universities in Turkey [16]. Duh et al. (2014) applied the DEA to measure performance with Taiwan colleges and universities as a research sample [17]. Nazarko and Šaparauskas (2014) conducted efficiency evaluation of Polish universities of technology by DEA [18]. Ding and Zeng (2015) applied TOPSIS and entropy weight methods to evaluate the 68 universities in China [19]. Dai and Li (2016) calculated the indicators' weight by AHP and rank the private higher education by fuzzy comprehensive assessment method [20]. Yang et al. (2018) constructed a two-stage network model and used Luenberger indicators to analyze Chinese research universities' productivity and evolution process [21]. Furthermore, whether the study results are meaningful relies on the reliability and validity of the evaluation. Nazari-shirkouhi et al. (2020) provided a comprehensive performance evaluation framework for performance assessment of the non-profit universities by integrating ANP-DEMATE-IPA technique in the fuzzy environment based on BSC approach [6]. Xu et al. (2022) established a comprehensive evaluation index system by the AHP-FCE method and applied the ARMA time series model to predict the development trend of Japanese higher education [22]. Texeira-Quiros et al. (2022) applied multiple linear regression to analyze organizational performance of higher education institutes [23]. From the aforementioned articles, the method of evaluation in higher education varies from statistics to MCDM. More specifically, statistical analysis cannot cover more evaluation criteria, and DEA only focus on financial criteria. Moreover, many methods of MCDM can be interpreted in terms of value judgments. Improper human judgments raise the level of vagueness in actual decision-making situations [24]. Furthermore, EWM is currently used as the most common objective method [25]. And, GRA clearly outperforms the other MCDM methods which prove its universal applicability and flexibility as an effective MCDM tool in solving complex decision-making problems [26]. Its' major advantages are based on original data, easy calculations and being straightforward and one of the best methods to decide in business environment [27]. Simultaneously, the application of EWM and GRA in educational evaluation is significantly less. GRA is broadly applied in evaluating or judging the performance of a complex project with meager information [28]. Thus, these provide broad space for the further exploration of the study on the performance measurement of universities.

In terms of evaluation criteria, the studies emphasize the student-oriented and timeliness of evaluation [29–31]. As regards evaluation quality, the studies highlight the reliability and validity of an evaluation [32,33].

2.2. MCDM techniques

Until 2019, 56 single and mixed MCDM methods were reported [34]. Each method has been developed with different advantages and disadvantages, thus, the scholars usually choose a technique according to the nature and intricacy of the problem [35].

FDM is a combination of the traditional Delphi method with fuzzy theory to solve some of the ambiguity of the expert consensus [36]. Since main focus is uncertainty and linguistic variables, the method utilizes triangular fuzzy number to determine the distance between the levels of consensus from the expert panel. Taking into account fact that it is less complicated and less time-consuming, the study intends to determine the suitability of indicators through FDM based on evaluation framework in analysis report of Zhejiang Higher Vocational Colleges.

As introduced by Shannon (1948), the entropy is a measure of uncertainty in information [37]. The lower the entropy of the criterion, the more valuable information the criterion contains [38], and vice versa. It explains the relative intensities of the criterion importance depend on the discrimination among data to evaluate the relative weights [39]. In light of avoiding human interference as well as enhancing objectivity, EWM has been widely utilized in many fields, including Cyber Security [40], Water Resource Management [41], Coal Mines [42], Power Quality [43], Energy Management [44], Supply Chain Management [45], etc. Hence, EWM was considered to be suitable for all the decision-making processes that require weight determination [25].

As an important branch of grey system theory, GRA is an effective method to not only cater to the context of poor, incomplete and uncertain information but also solve complex interrelationships among multiple variables and factors [27]. GRA uses the relational grade to measure the similarity between the comparability and reference sequences [46]. The highest grey relational grade among the alternatives will be the best choice [47]. Namely, it is tight to determine whether the relation between different series based on the geometry of the series curve [48]. Multi criteria problems can be effectively handled using GRA [49], and it assists ranking considering all quantitative criteria [50]. As it gradually matured, GRA has been widely applied as MCDM method to solve problems in many fields, such as international trade [51], project management [52], financial performance assessment [53], healthcare service [54], electricity power generation [55], quality of fruit [56], efficiency of public hospitals [57], etc.

Performance measurement is a term which contains program input, output, intermediate outcomes and end outcomes [58]. In this study, the data belongs to second-hand and quantitative data of staged development about higher vocational colleges. Given the long-term characteristic and the complexity of identification for the output variables of the educational institutions, reaching a broad consensus on determination of output and input variables in a short term is difficult. Because EWM and GRA are categorized into objective method, comparing to subjective methods, objective methods combine the strength comparison of each criterion with the conflict between the criteria [59]. Thus, objective methods demonstrate the greater advantages to avoid the influence of subjective assessment and improve the accuracy and reliability of analysis.

3. Methods and material

3.1. Research design

This study proposes a comprehensive evaluation model combining FDM, EWM, and GRA to assess college performance, and tests the difference of results from different methods. Compared with prior studies, the major difference of the method is to greatly reduce the controversy about attributes of factors and reasonably distribute weights through adequate utilization of data by objective weight methods. More important than the definition of models is its effective implementation, use and revision [12]. In brief, FDM is used to screen the existing performance factors; then, the weights of criteria are calculated by the EWM and GRA; Furthermore, grey relational degrees are calculated by multiplying the grey relational coefficients by criteria' weights; Ranking the alternatives and performing difference test about variables are the final steps. Fig. 1 depicts an illustration of the proposed method and the calculation steps.

3.2. Data source and evaluation framework

From 1980 to 2010, about 30 models of organizational performance measurement system have emerged [60]. No matter which aspect we measure, building a robust evaluation framework is essentially a multiple criteria problem [61]. Furthermore, strengths and weaknesses of universities can be recognized by performance evaluation systems [62].

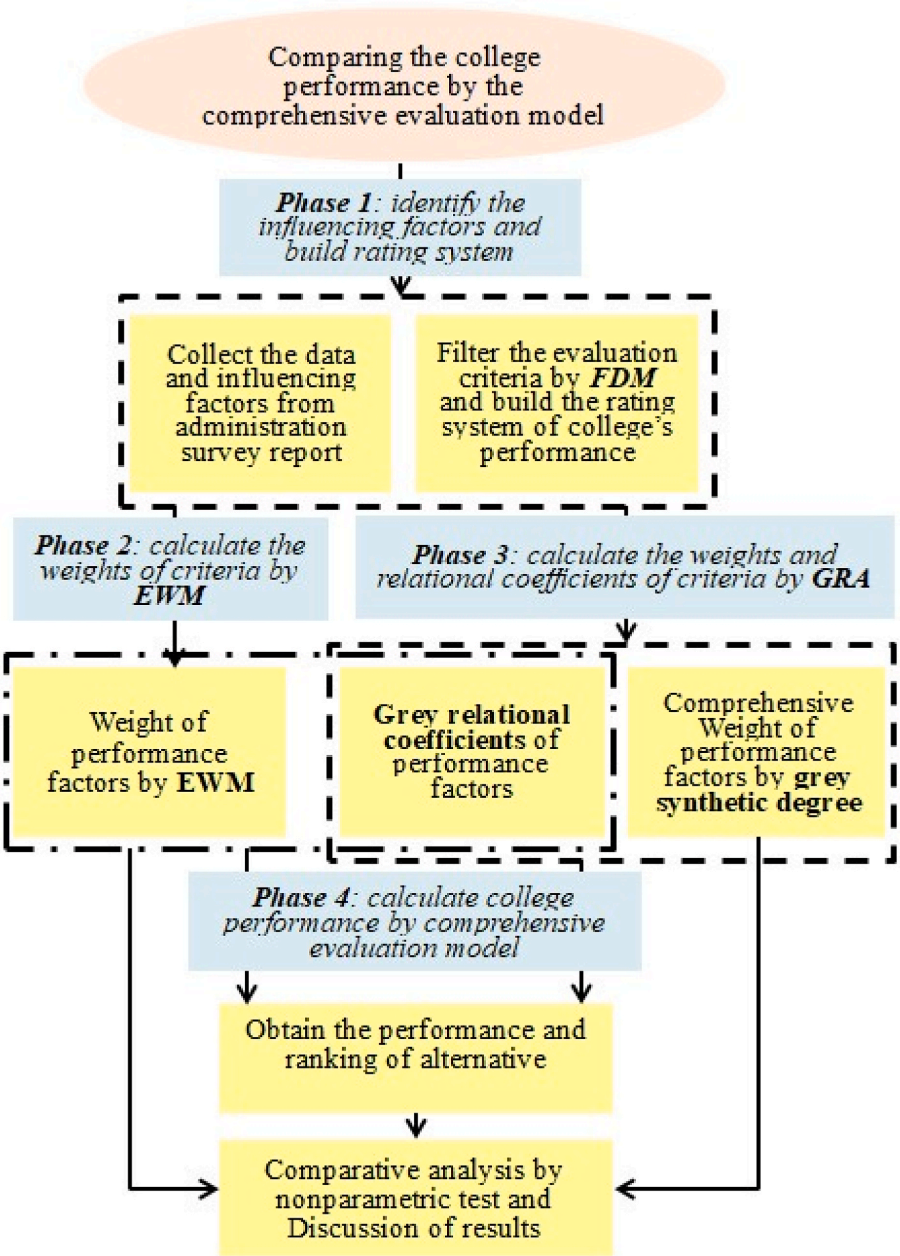

Fig. 1. Analysis process.

The data comes from the "Analysis Report on the Status of Talent Cultivation in Zhejiang Higher Vocational Colleges in 2014–2015 Academic Year", which covers 49 factors of talent cultivation in 46 higher vocational colleges in Zhejiang Province during the school-running process, including student-teacher ratio (A01), paper books per student (A02), teaching and administrative space per student (A03), proportion of full-time teachers with postgraduate degrees (A04), value of teaching and scientific research equipment per student (A05), average occupied space per student (A06), average dormitory area per student (A07), ratio of teachers with senior professional titles to full-time teachers (A08), teaching computers for 100 students (A09), average annual number of books bought per student (A10), proportion of newly increased teaching and research equipment (A11), income per student (A12), tuition income (A13), financial subsidies (A14), special investment (A15), social donations (A16), other income (A17), land acquisition (A18), infrastructure construction (A19), equipment procurement (A20), daily teaching (A21), teaching reform (A22), teacher construction (A23), book purchase (A24), other expenditures (A25), student-teacher ratio of full-time teachers (A26), the number of full-time teachers per (enrollment) major (A27), the proportion of full-time teachers under the age of 35 (A28), the proportion of full-time teachers with doctoral/master degrees (A29), the proportion of the full-time teachers with double-qualified teachers (A30), award-winning projects per 100 full-time teachers (A31), technology patents per 100 full-time teachers (A32), per capita payment quota for research projects hosted by full-time teachers (A33), the number of published papers, works and publications per 100 full-time teachers (A34), per capita training days of full-time teachers (A35), per capita number of temporary training days of full-time teachers (A36), per capita social part-time days of full-time teachers (A37), per capita teaching construction's expenses of full-time teachers (A38), number of off-campus part-time teachers hired by majors (A39), proportion of off-campus part-time teachers (A40), proportion of off-campus part-time teachers with junior/senior professional and technical positions (A41), expenses for hiring off-campus part-time teachers for each major (A42), the area of practice sites per student (A43), total value of practice base equipment per student (A44), the ratio of newly added equipment in the current year to the total value of equipment (A45), cost of raw materials (consumables) per student (A46), cost of equipment maintenance per student (A47), frequency of students' use by in the on-campus training base (A48), allocation ratio of the number of workstations to training items (A49).

Given the indicators above in the report, A1–A5 were grouped into basic indicators of college, A6–A11 were classified into monitoring indicators of college. While financial indicators A12–A25 covered income and expenditure of college and A26–A42 focused on the construction of teacher team in vocational colleges. Finally, A43–A49 served as indicators of college's practice base. Therefore, these indicators revealed the performance level of Zhejiang vocational colleges to a certain extent.

3.3. Proposed combinations of FDM, entropy and GRA methods

3.3.1. Fuzzy Delphi method

FDM can be summarized in the following steps.

Step 1: Scoring the factors.

After designing an FDM questionnaire for the evaluation factors and forming an appropriate panel, each expert must give three values for each factor, including the minimum value, the maximum value and the optimal value (A). Wherein, the minimum value means the most conservative perception value of the factors by experts, the maximum value represents the most optimistic perception value of the factors by experts, the optimal value (A) of each factor is the intuitive value of the individual's importance to each factor given by experts.

Step 2: Forming the triangular fuzzy number of the factors.

For each evaluation factor, the counting of the most conservative and most optimistic values given by the panel, and the extreme values that fall outside twice the standard deviation are deleted. Then, the maximum, minimum and geometric mean values for the remaining most conservative and most optimistic values that have not been excluded are calculated. Respectively, the most conser-

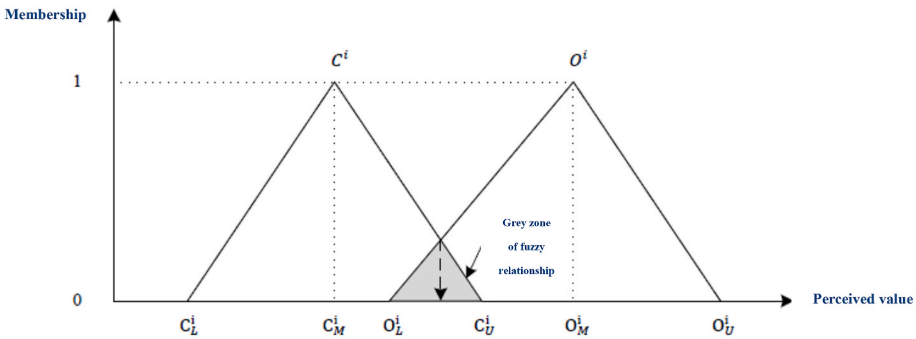

Fig. 2. Grey zone diagram of the fuzzy relationship between the two triangle fuzzy numbers.

vative triangular fuzzy number $C^i = (C^i_L, C^i_M, C^i_U)$ and most optimistic triangular fuzzy number $O^i = (O^i_L, O^i_M, O^i_U)$ are calculated for each evaluation factor as shown in Fig. 2:

Step 3: Judging the consensus of expert opinions by the following rules.

1. If no overlap occurs between the two triangular fuzzy numbers, that is $(C^i_U \le O^i_L)$, then the opinions of each expert have a consensus section according to the interval value. Therefore, the consensus value G^i of the evaluated factor is equal to the arithmetic average value of C^i_M and O^i_M, as shown in Eq. (1):

$$G^i = \frac{(C^i_M + O^i_M)}{2} \tag{1}$$

2. When the overlap emerges between the two triangular fuzzy numbers, that is $(C^i_U > O^i_L)$, and the grey area of the fuzzy relationship $Z^i = C^i_U - O^i_L$ is less than the interval range $M^i = O^i_M - C^i_M$ of the geometric average value of the optimistic cognition and the conservative cognition of the evaluation factor, then the opinions of each expert does not reach a consensus as a whole. The opinions that experts give extreme scores do not have too much difference from the opinions of the others, so the result does not lead to disagreement. Therefore, the factors consensus importance value G^i is calculated as shown in Eq. (2(2) and (3)(3)

$$F^i(X_j) = \left\{ \int x\{\min[C^i(X_j), O^i(X_j)]\}dx \right\} \tag{2}$$

$$G^i = \{X_j | \max \mu_{F^i}(X_j)\} \tag{3}$$

3. When the overlap emerges between the two triangular fuzzy numbers, that is $(C^i_U > O^i_L)$, but the grey area of the fuzzy relationship $Z^i = C^i_U - O^i_L$ is more than the interval range $M^i = O^i_M - C^i_M$ of the geometric average value of the optimistic cognition and the conservative cognition of the evaluation factor, then the opinion interval value of each expert has no consensus section, and too different scores of the panel lead to divergence of opinion. Steps 1 to 3 must be repeated until all evaluation factors reach convergence and the consensus importance value G^i is obtained.

Step 4: Setting a threshold value to construct the evaluation system.

The selection of the evaluation criteria is based on the comparison of the threshold value S and the consensus importance value G^i, and the threshold value setting standard directly affects the number of screening evaluation criteria. According to prior papers, the method of threshold value setting includes the threshold value range of approximately between 6 and 7 [63], subjective identification of decision makers [64], arithmetic average [65].

3.3.2. *Entropy weight method*

In the method, m criteria and n samples are set in the evaluation, and V_{ij} denotes the performance value of the kth attribute of the ith alternative. First of all, it is to standardize the data, now that these data are not uniform and can't be processed. Thus, the data in the decision matrix is normalized according to Eqs. (4) and (5):

$$X'_{ki} = (V_{ki} - \min(V_i)) / (max(V_i) - \min(V_i)) \tag{4}$$

$$X'_{ki} = (max(V_i) - V_{ki}) / (max(V_i) - \min(V_i)) \tag{5}$$

The entropy value E_k of the jth criterion is defined as Eq. (6):

$$E_k = -K * \sum_{k=1}^{m} p_{ki} * \ln p_{ki} \tag{6}$$

wherein, $p_{ki} = X'_{ki} / \sum_{k=1}^{n} X'_{ki}$, and K is a constant of the entropy equation, that is, $K = 1/\ln n$.

Finally, the indicators' weights, w^1_k is obtained by the following expression:

$$w^1_k = (1 - E_k) \Big/ \sum_{k=1}^{m} (1 - E_k) \tag{7}$$

where $(1 - E_k)$ reflects the diversity degree that the outcomes of the kth criterion involves information. Thus, because the entropy value means the degree of disorder in the system, the lower entropy value, the higher weight [44].

3.3.3. *Grey relational analysis*

The major steps of GRA are as follows:

Step 1: Normalizing data.

According to the target of the indicator (Eqs. (4) and (5)), we normalize the original evaluation matrix, so the normalized decision matrix T is obtained:

$$T = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \cdots & \gamma_{mn} \end{bmatrix} \tag{8}$$

Step 2: Calculating the grey relational coefficient.

The correlation coefficient formula is a function of the distance between the reference and comparison sequences [66]. For sequence comparison, selecting a reference sequence is a key link. Generally speaking, the traditional method chooses the maximum value of the positive indicator and the minimum of the negative indicator, or the average value. The reference sequence γ_{k0} gets from the best indicator of alternative from normalization matrix. With that, other sequence γ_{ki} acts as a comparative sequence. Then the globalized grey relational coefficient is defined as:

$$\delta_{ki} = \frac{\Delta_{min} + \zeta\,\Delta_{max}}{\Delta_{ki} + \zeta\,\Delta_{max}} \tag{9}$$

where, $\Delta_{ki} = |\gamma_{k0} - \gamma_{ki}|$, $\Delta_{min} = \min_k \min_i |\gamma_{k0} - \gamma_{ki}|$, $\Delta_{max} = \max_k \max_i |\gamma_{k0} - \gamma_{ki}|$, ζ is the distinguishing coefficient, and $\zeta \in [0, 1]$. In general, the distinguishing coefficient is set as 0.5 [67], because of offering moderate distinguishing effects and good stability.

Step 3 : Determining the grey relational matrix.

The grey relational matrix E can be established as follows.

$$E = \begin{bmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1n} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \delta_{m1} & \delta_{m2} & \cdots & \delta_{mn} \end{bmatrix} \tag{10}$$

Step 4: Calculating different grey relational degree by different methods.

ε_{0i} is calculated as follows by Eq. (11), which is absolute relational degree (ARD);

$$\varepsilon_{0k} = \frac{1 + |s_0| + |s_k|}{1 + |s_0| + |s_k| + |s_k - s_0|} \tag{11}$$

wherein $|s_0|$ is the absolute value of sum of the normalized values of reference series, $|s_k|$ is the absolute value of sum of the initial values.

$$|s_0| = \left|\sum_{k=2}^{m-1} x_0^0(k) + \frac{1}{2}x_0^0(n)\right|,\ |s_k| = \left|\sum_{k=2}^{m-1} x_k^0(k) + \frac{1}{2}x_k^0(n)\right|,$$

$$|s_k - s_0| = \left|\sum_{k=2}^{m-1}\left(x_k^0(k) - x_0^0(k)\right) + 1 \Big/ 2\left(x_k^0(n) - x_0^0(n)\right)\right| \tag{12}$$

η_{0i} is calculated as follows by Eq. (14), which is relative relational degree (RRD).

$$\eta_{0k} = \frac{1 + |s_0^{'}| + |s_k^{'}|}{1 + |s_0^{'}| + |s_k^{'}| + |s_k^{'} - s_0^{'}|} \tag{13}$$

$$|s_0^{'}| = \left|\sum_{k=2}^{m-1} x'^0_0(k) + \frac{1}{2}x'^0_0(n)\right|,\ |s_k^{'}| = \left|\sum_{k=2}^{m-1} x'^0_k(k) + \frac{1}{2}x'^0_k(n)\right|$$

$$|s_k^{'} - s_0^{'}| = \left|\sum_{k=2}^{m-1}\left(x'^0_k(k) - x'^0_0(k)\right) + 1 \Big/ 2\left(x'^0_k(n) - x'^0_0(n)\right)\right| \tag{14}$$

Step 5: Calculating synthetic weight by synthetic relational degree (SRD).

According to Luo (2005) [68], synthetic relational degree ρ_{0k} is calculated as follows:

$$\rho_{0k} = \theta\varepsilon_{0k} + (1-\theta)\eta_{0k}, \theta \in [0,1] \tag{15}$$

where θ is taken as 0.5 [69], that means that both are equally important.

With that, each indicator's weight will be calculated in term of Eq. (16) from the synthetic relational degree.

$$w_k^2 = \frac{\rho_{0k}}{\rho_{01} + \rho_{02} + \ldots + \rho_{0p}}, k = 1, 2, \ldots, m, \tag{16}$$

Ultimately, the weight vector $W^2 = [w_1^2, w_2^2, \ldots w_m^2]$ is obtained.

Step 6 : Calculating performance by comprehensive evaluation model

Relational degree is calculated by relational coefficient matrix E multiplied by the weight vector W^1 or W^2 as follows:

$$\mathrm{R} = (r_1, r_2, \ldots, r_n) = W * E \tag{17}$$

By sorting by the relational degree r_i, the ranking of the relational degree indicates the ranking of the comprehensive evaluation strength of each alternative.

4. Empirical analysis

4.1. Factors screening

With regards to the size of experts of the panel, no rule for determining the appropriate panel's size has been established [70]. Skulmoski et al. (2007) suggested that a smaller panel should consist of 10–15 persons in a homogeneous group [71]. Hence, in light of professionalism of educational evaluation, 10 questionnaires of FDM were issued and retrieved via email and face-to-face. Table 1 demonstrates the characteristic of respondents.

Meanwhile, the criteria in Section 3.1 consisted of the questionnaire of FDM, in which a 0–10 rating scale was adopted, and the higher the score, the more important it is. This study adopted arithmetic mean to set the threshold value. Specifically speaking, the overall arithmetic mean was calculated based on the geometric mean of the optimal value of the panel, and then, the threshold value was equivalent to the arithmetic mean multiplied by 0.8 in the light of the 80/20 rule. Finally, the threshold value of this study was 4.328.

According to the calculation steps of the aforementioned FDM, opinions of the panel were sorted up in Table 2:

The indicators, including A10, A17, A18, A20, A24, A25, A31, A32, A33, A34, A37, A39, A40, A42, A46, and A47, were not adopted in the consequent study because their consensus value was less than the threshold value (4.328). This implies that there was no consensus in the panel on these indicators. Moreover, the remaining 33 indicators constituted the evaluation system of the higher vocational colleges.

4.2. Weights obtained from entropy weight methods and GRA

In term of Eqs. (4)–(4)–(7), the Entropy weights of indicators were calculated, furthermore, according to Eqs. (8)–(8)–(16), the absolute relational degree (ARD), relative relational degree (RRD), synthetic relational degree (SRD) and indicators' synthetic weights were obtained, which are all shown in Table 3.

As shown in the table above, the top 3 indicators of synthetic weights were A35 (0.0352), A22 (0.0345) and A23 (0.0337), however,

Table 1
Information of respondents.

No.	Gender	Working time (year)	Education degree	Region
1	Male	28	Doctor	Hangzhou
2	Male	20	Bachelor	Hangzhou
3	Female	33	Master	Hangzhou
4	Female	16	Doctor	Hangzhou
5	Male	22	Doctor	Hangzhou
6	Female	18	Master	Jinhua
7	Male	21	Doctor	Huzhou
8	Female	19	Doctor	Jiaxing
9	Female	13	Bachelor	Ningbo
10	Male	25	Bachelor	Ningbo

Table 2
Summarization result of Fuzzy Delphi Method.

Criteria	C		O		A		Geometric Mean			M	Z	Verification	Expert Consensus G
	Min	Max	Min	Max	Min	Max	C	A	O			M-Z	
A01	5	8	9	10	8	10	6.903	8.565	9.791	2.888	1	1.888	8.347
A02	2	4	4	9	3	6	2.862	4.227	5.966	3.104	0	3.104	4.414
A03	4	7	6	10	5	8	4.891	6.093	7.432	2.541	−1	3.541	6.405
A04	2	8	4	10	3	10	4.373	6.093	7.591	3.217	−4	7.217	5.990
A05	3	7	6	10	5	9	4.618	6.239	7.873	3.255	−1	4.255	6.440
A06	3	6	6	10	5	8	4.441	6.258	7.689	3.248	0	3.248	6.065
A07	3	7	6	10	5	8	5.008	6.491	7.873	2.865	−1	3.865	6.485
A08	2	5	4	10	3	8	3.758	5.502	6.903	3.145	−1	4.145	4.700
A09	2	5	5	8	3	6	3.438	4.856	6.518	3.080	0	3.080	4.978
A10	1	4	3	8	2	6	2.352	3.728	4.919	2.567	−1	3.567	3.538
A11	1	5	4	9	3	7	2.825	4.618	5.966	3.141	−1	4.141	4.475
A12	4	8	6	10	5	10	5.448	7.414	8.841	3.392	−2	5.392	7.054
A13	3	8	6	10	5	10	4.919	6.749	8.407	3.488	−2	5.488	6.877
A14	5	8	8	10	7	9	5.908	7.560	9.364	3.456	0	3.456	8.000
A15	5	7	8	10	6	8	5.547	7.354	9.169	3.622	1	2.622	7.358
A16	2	4	4	8	3	5	2.702	4.129	5.985	3.283	0	3.283	4.344
A17	0	3	4	7	2	4	0.000	3.104	4.891	4.891	1	3.891	2.446
A18	0	4	3	8	2	5	0.000	3.594	5.335	5.335	−1	6.335	3.369
A19	0	7	8	10	5	8	0.000	6.722	8.769	8.769	1	7.769	7.901
A20	0	4	4	8	3	6	0.000	4.682	6.402	6.402	0	6.402	3.201
A21	0	8	4	10	3	9	0.000	5.966	7.796	7.796	−4	11.796	5.287
A22	4	5	6	10	5	8	4.373	5.908	7.591	3.217	1	2.217	5.982
A23	4	8	8	10	7	9	5.448	7.560	9.364	3.916	0	3.916	7.406
A24	2	3	4	9	3	6	2.352	4.282	5.753	3.400	1	2.400	4.052
A25	0	3	4	5	2	4	0.000	2.702	4.573	4.573	1	3.573	2.287
A26	5	8	9	10	7	10	5.875	8.145	9.587	3.712	1	2.712	7.731
A27	5	8	9	10	7	10	5.875	8.145	9.587	3.712	1	2.712	7.731
A28	2	5	4	10	3	7	2.993	4.618	6.093	3.101	−1	4.101	4.510
A29	2	5	5	10	4	8	3.245	5.253	6.815	3.570	0	3.570	5.030
A30	3	8	8	10	6	10	5.448	7.665	9.364	3.916	0	3.916	7.406
A31	1	3	3	8	2	5	1.644	3.245	4.743	3.099	0	3.099	3.193
A32	1	5	3	9	2	6	2.268	4.043	5.387	3.119	−2	5.119	3.933
A33	1	3	3	9	2	6	2.048	3.898	5.387	3.339	0	3.339	3.717
A34	1	3	3	10	2	7	2.221	3.845	5.144	2.923	0	2.923	3.682
A35	0	7	3	9	2	8	0.000	4.891	6.355	6.355	−4	10.355	4.296
A36	1	8	4	10	3	9	3.314	5.809	7.456	4.141	−4	8.141	5.698
A37	0	5	3	9	2	7	0.000	4.618	6.355	6.355	−2	8.355	3.803
A38	2	8	6	10	4	9	3.949	6.188	8.278	4.330	−2	6.330	6.720
A39	1	5	3	9	2	7	2.268	4.939	6.507	4.239	−2	6.239	4.124
A40	1	5	3	9	2	7	2.091	4.258	5.785	3.694	−2	5.694	3.978
A41	1	7	3	9	2	8	1.838	4.282	5.842	4.003	−4	8.003	4.420
A42	1	5	3	9	2	7	1.974	4.020	5.533	3.558	−2	5.558	3.911
A43	2	7	7	9	5	8	4.020	6.320	8.139	4.119	0	4.119	6.080
A44	1	8	5	10	4	9	3.594	5.966	7.591	3.996	−3	6.996	6.111
A45	1	5	5	9	4	6	2.268	5.144	6.470	4.202	0	4.202	4.369
A46	0	4	3	8	2	6	0.000	3.594	5.502	5.502	−1	6.502	3.385
A47	0	4	3	7	2	6	0.000	3.438	5.008	5.008	−1	6.008	3.334
A48	1	8	5	10	4	9	3.594	6.153	7.982	4.388	−3	7.388	6.211
A49	1	7	7	10	4	8	4.020	6.207	8.119	4.099	0	4.099	6.070

the top 3 indicators of Entropy weights were A43 (0.1072), A16 (0.1016) and A12 (0.0925). Under synthetic weights, the weight's range was in the interval [0.0239, 0.0352], on the contrary, the weight's range was in the interval [0.0034, 0.1072] under Entropy weights. Hence, this means that the weights vary more dramatically on the entropy method than on the grey synthetic weight method. Moreover, from the perspective of the indicators weight, the two methods highlighted financial indicators more. EWM also focused on the indicators of practice base. More importantly, the construction indicators of teacher team, which were ignored in EWM, ranked 1st in GRA as a whole. Additionally, GRA still underlined the monitoring indicators.

Besides, different weights of indicators were performed using nonparametric Wilcoxon signed ranks test in order to inspect the difference between them as followed in Table 4:

According to the test results in table above, statistical test Z was −1.010, and two-tailed P value (0.313) obviously exceeded 0.05 under a significant level of 0.05. Therefore, the null hypothesis needs to be kept, namely, there was no significant difference between the two types of weights.

Table 3
Indicators' weights and relational degrees.

Indicator	ARD	RRD	SRD	Synthetic Weight	Entropy Weight
A01	0.5615	0.997	0.7792	0.0292	0.0206
A02	0.6620	0.8480	0.7550	0.0283	0.0211
A03	0.6181	0.9327	0.7754	0.0291	0.0546
A04	0.6092	0.9096	0.7594	0.0285	**0.0034**
A05	0.6556	0.8822	0.7689	0.0288	0.0443
A06	0.5679	0.9504	0.7592	0.0285	0.0264
A07	0.5958	0.9747	0.7853	0.0294	0.0376
A08	0.5404	0.8336	0.6870	0.0257	0.0076
A09	0.6221	0.8706	0.7464	0.0280	0.0141
A11	0.8224	0.8968	0.8596	0.0322	0.0148
A12	0.5992	0.9661	0.7827	0.0293	0.0925
A13	0.5110	0.7639	0.6375	**0.0239**	0.0127
A14	0.9357	0.8867	0.9112	0.0342	0.0200
A15	0.6131	0.8875	0.7503	0.0281	0.0580
A16	0.7070	0.8905	0.7988	0.0299	0.1016
A19	0.8241	0.9012	0.8626	0.0323	0.0567
A21	0.8144	0.9012	0.8578	0.0321	0.0170
A22	0.9412	0.9005	0.9209	0.0345	0.0201
A23	0.8946	0.9012	0.8979	0.0337	0.0194
A26	0.7639	0.8715	0.8177	0.0306	0.0052
A27	0.8825	0.9026	0.8926	0.0335	0.0149
A28	0.7824	0.9035	0.8429	0.0316	0.0096
A29	0.7271	0.9000	0.8136	0.0305	0.0080
A30	0.8979	0.8982	0.8981	0.0337	0.0134
A35	0.9782	0.8982	0.9382	0.0352	0.0104
A36	0.7827	0.8933	0.8380	0.0314	0.0120
A38	0.5142	0.9990	0.7566	0.0284	0.0386
A41	0.8497	0.8947	0.8722	0.0327	0.0114
A43	0.563	0.8888	0.7259	0.0272	0.1072
A44	0.6725	0.8959	0.7842	0.0294	0.0253
A45	0.7141	0.8924	0.8032	0.0301	0.0267
A48	0.6626	0.8891	0.7758	0.0291	0.0434
A49	0.7531	0.9021	0.8276	0.0310	0.0315

Table 4
Wilcoxon signed rank test of indicators' weights.

Weight methods	Level	Number	*Z*	*P* value	Decision
Synthetic weight - Entropy weight	Negative	11[a]	−1.010[a]	0.313	Keep the null hypothesis with no difference
	Positive	22[b]			
	Zero	0[c]			
	Total	33			

[a] Synthetic weight < Entropy weight;
[b] Synthetic weight > Entropy weight;
[c] Synthetic weight = Entropy weight.

4.3. Relational degree obtained by different methods

The relational degrees of alternatives can be calculated according to the relational coefficient in conjunction with different weights of indicators, and ranked as shown in Table 5.

As can be seen from the table above, under Entropy weight-based relational degree, value's range was in the interval [0.8606, 1.0000], meanwhile, under synthetic weight-based relational degree, value's range was in the interval [0.9003, 1.0000]. Specifically speaking, the top 3 samples of Entropy weight-based relational degree were 1(1.0000), 21(0.9933) and 16 (0.9915), whereas under synthetic weight-based relational degree, the top 3 samples were 1(1.0000), 9(0.9894) and 21 (0.9884). Furthermore, the last 5 alternatives under both methods were exactly identical, namely 41(0.8606, 0.9003), 30(0.9440, 0.9466), 46(0.9496, 0.9490), 43 (0.9633, 0.9548) and 4 (0.9652, 0.9586), respectively.

With that, different relational degrees of indicators were performed using nonparametric Wilcoxon signed ranks test in order to examine the difference between them as shown in Table 6 below:

According to the above-mentioned test results, statistical test $\boldsymbol{Z}$ equals to −4.856, and two-tailed P value (0.000) was obviously less than 0.05 under a significant level of 0.05. As a result, the null hypothesis needs to be rejected, which implies that there was a significant difference between the two relational degrees.

Likewise, Kendall's W test was used to test whether there was a significant difference in the rankings of relational degrees based on

Table 5
Relational degrees and ranking of the alternative.

Sample	Relational degree - entropy	Rank	Relational degree -grey	Rank
1	1.0000	1	1.0000	1
2	0.9899	7	0.9873	7
3	0.9882	14	0.9835	13
4	0.9652	**42**	0.9586	**42**
5	0.9840	23	0.9815	19
6	0.9840	24	0.9802	22
7	0.9696	39	0.9744	35
8	0.9837	26	0.9817	18
9	0.9909	5	0.9894	2
10	0.9875	15	0.9831	14
11	0.9897	9	0.9876	6
12	0.9823	29	0.9776	29
13	0.9693	40	0.9668	41
14	0.9883	13	0.9848	12
15	0.9838	25	0.9809	20
16	0.9915	3	0.9879	4
17	0.9855	17	0.9829	15
18	0.9847	22	0.9806	21
19	0.9767	37	0.9682	40
20	0.9899	8	0.9863	9
21	0.9933	2	0.9884	3
22	0.9809	31	0.9795	25
23	0.9753	38	0.9727	37
24	0.9907	6	0.9865	8
25	0.9851	21	0.9780	27
26	0.9801	33	0.9764	33
27	0.9913	4	0.9854	11
28	0.9797	34	0.9754	34
29	0.9893	11	0.9855	10
30	0.9440	**45**	0.9466	**45**
31	0.9680	41	0.9690	38
32	0.9894	10	0.9878	5
33	0.9855	18	0.9772	31
34	0.9773	36	0.9684	39
35	0.9835	27	0.9773	30
36	0.9791	35	0.9786	26
37	0.9805	32	0.9737	36
38	0.9888	12	0.9829	16
39	0.9858	16	0.9822	17
40	0.9852	20	0.9797	24
41	0.8606	**46**	0.9003	**46**
42	0.9814	30	0.9776	28
43	0.9633	**43**	0.9548	**43**
44	0.9831	28	0.9764	32
45	0.9855	19	0.9801	23
46	0.9496	**44**	0.9490	**44**

Table 6
Wilcoxon signed ranks test of alternatives' relational degrees.

Weight methods	Level	Number	Z	P value	Decision
Grey comprehensive weight-based relational degree: Entropy weight-based relational degree	Negative	42[a]	−4.856[b]	0.000	Reject the null hypothesis with difference
	Positive	4[b]			
	Zero	0[c]			
	Total	46			

[a] Grey comprehensive weight-based relational degree < Entropy weight-based relational degree;
[b] Grey comprehensive weight-based relational degree > Entropy weight-based relational degree;
[c] Grey comprehensive weight-based relational degree = Entropy weight-based relational degree.

the two weights as shown in Table 7.

As indicated by the figure in the table, P value (0.739) of the Kendall's W test about two relational degrees' ranking evidently surpassed 0.05 under a significant level of 0.05. Consequently, the null hypothesis needs to be maintained, which indicates that there was no significant difference in the ranking of relational degrees.

Table 7
Kendall's W test about different ranking.

Rank methods	Number	Average grade	Kendall's *W* coefficient	*P* value	Decision
Grey comprehensive weight-based relational degree	46	1.48	0.002	0.739	Keep the null hypothesis with no difference
Entropy weight-based relational degree	46	1.52			

4.4. Difference test about different types of relational degree

In order to better understand the differences in the performance of specific types about higher vocational colleges in Zhejiang, this study conducted a more in-depth discussion by dividing 46 higher vocational colleges into public colleges and private colleges in term of business entities, and colleges in Hangzhou and in non-Hangzhou in term of geographical locations. Moreover, the Mann-Whitney U method was used to test whether there were differences in school performance as is shown in Table 8:

As can be seen from the table above, under grey comprehensive weight-based relational degree, there was a significant difference between the relational degrees of public colleges and private colleges because *P* value (0.048) was slightly lower than 0.05 under a significant level 0.05. On the contrary, there was no significant difference between the relational degrees of colleges in Hangzhou and non-Hangzhou, because *P* value (0.352) was greater than 0.05. Meanwhile, under Entropy weight-based relational degree, no significant difference existed exactly at the business entities as well as geographical locations with *P* values (0.146, 0.163) respectively, which both were greater than 0.05.

5. Discussion

This paper proposed a hybrid model to conduct the performance measurement and comparison about the relational degrees of higher vocational college in Zhejiang. First of all, FDM filtered out the indicators in evaluation framework; Secondly, Entropy weight method was employed for the indicators' weight; Next, GRA was implemented to obtain the grey correlation coefficient and the indicator synthetic weights; Furthermore, the relational degrees of the alternatives was obtained and ranked through comprehensive measurement model; Finally, a non-parametric test was carried out on the indicators' weights, relation degrees and ranking; and then, according to the business entity and geographical location of relational degrees, the Mann-Whitney *U* test was used to examined difference features of relational degrees between different groups.

The analysis of expert opinions by FDM resulted in the rejection of 16 indicators, which experts did not reach a consensus. In turn, the remaining 33 indicators served as the foundation for subsequent study and evaluation.

The core of the comprehensive measurement model is to carefully choose the calculation method to determine the weight of each indicator. Different weight calculation methods will lead to inconsistent measurement results. Therefore, whether or not the weight determination is reasonable will directly affect the reliability and validity of the decision results. Although the weights varied more greatly on EWM than on GRA, no significant difference was detected among them using a nonparametric test. Specially speaking, financial indicators were emphasized in both methods, this means that higher vocational colleges in Zhejiang were experiencing a boom in term of increasing investment and construction. Notwithstanding, from the analytical results, EWM focused more on hardware construction (practice base), in turn, GRA further highlighted the software construction (construction of teacher team).

Furthermore, based on the weights mentioned previously, relational degrees of the alternatives were calculated by comprehensive measurement model, at the same time, significant difference did exist between the two relational degrees. Hence, the difference of relational degrees originates possibly from difference in the internal mechanism of data analysis between Entropy weight method and synthetic weight method of GRA. On the contrary, there was no significant difference between rankings of the alternatives according to relational degrees. This means that the difference of relational degrees had little influence on the rankings of the alternatives. Also, the

Table 8
Mann-Whitney *U* test about different types of different relational degree.

Relational degree	Type		Number	Average	*P* value	Decision
Grey comprehensive weight-based relational degree	Business entities	Public colleges	37	0.9764	0.048	Reject the null hypothesis with difference
		Private colleges	9	0.9734		
	geographical locations	Colleges in Hangzhou	20	0.9756	0.352	Keep the null hypothesis with no difference
		Colleges in non-Hangzhou	26	0.9763		
Entropy weight-based relational degree	Business entities	Public colleges	37	0.9729	0.146	Keep the null hypothesis with no difference
		Private colleges	9	0.9762		
	geographical locations	Colleges in Hangzhou	20	0.9784	0.163	Keep the null hypothesis with no difference
		Colleges in non-Hangzhou	26	0.9791		

objective weight method optimizes results of the comprehensive measurement and improves the precision.

More than that, the study implemented a more in-depth comparative analysis of relational degrees in light of business entities and geographical locations. Regarding geographical location, there was no significant difference between two relational degrees about colleges in Hangzhou and non-Hangzhou, which indicates that geographical location imposed little impact on the colleges' performance. Higher vocational colleges should not only adapt measures to local conditions but also build their unique characteristics by relying on local and related industries. Thus, once these conditions are met, colleges in non-Hangzhou will be promising as well. Regarding business entity, the results of two relational degrees about public colleges and private colleges were not completely consistent. In general, the fact that public colleges had a higher overall ranking than private colleges reflects the difference to some extent. Combined with the actual situations, the performance disparity between public college and private colleges probably results from the policy and financial assistance of government. Government financial support was perceived as the most influential determinant for public institutions [72].

6. Conclusions

The present study bears certain significance with regard to examining colleges' performance and ranking the competitive colleges by combining FDM, EWM and GRA in China context. This study also enriches the ideas for conducting an objective and rational college evaluation using objective methods. FDM filtered the evaluation indicators based on the experts' assessment, so 33 indicators were selected as the base of the next stage. Subsequently, according to the weights, relational degrees and ranking, the study then conducted a comparative analysis. Although there was no significant difference between the weights from two methods, the relational degrees from two methods did contain significant difference. However, these did not exert any impact on the ranking of the alternatives. What' more, by the Mann-Whitney *U* test, based on geographical location of relational degrees, no significant difference was observed between colleges in Hangzhou and non-Hangzhou; on the contrary, controversial phenomenon emerged between public colleges and private colleges, which was divided by the relational degrees of business entity.

The study has significant implications for higher educational institutions. First, according to the overall ranking, higher educational institutions can clearly find out their position in Zhejiang. Compared with the top-ranked institutions, they are more likely to detect their own shortcomings and find the direction and breakthrough for subsequent development. Second, when the categories of performance indicators are unclear, applying objective methods to measure the performance should be an effective method. Meanwhile, in view of the context of the "*Shuang Gao*" of Chinese universities and colleges, this proposed model also becomes an effective tool for the administration institutions to rank universities and colleges. Third, since the two methods place different emphasis on the indicators' weights, administration institutions can also flexibly choose them according to the different focuses about performance measurement of college. Specifically speaking, EWM may be a better choice when the performance measurement concerns about college hardware construction, whereas GRA may be appropriate when the performance measurement highlights college software construction. Last but not least, comparative analysis of relational degrees about business entities and geographic locations greatly deepens the understanding of the characteristics of higher vocational education in Zhejiang. No difference of the rankings about colleges in Hangzhou and non-Hangzhou demonstrated the balance of the regional distribution of higher vocational education in Zhejiang, which was conducive to the balanced development of regional education. Additionally, the application of the weight method determined whether or not differences existed in business entities. Under EWM, there was no difference in the rankings of colleges of business entities from the angle of hardware. This implies that the public and private colleges were equally matched in term of performance measurement from the angle of hardware. On the contrary, under GRA, significant difference existed in the rankings of colleges of business entities from the angle of software (construction of teacher team). This means that the private colleges lagged far behind the public colleges in term of performance measurement from the angle of software (construction of teacher team).

Some limitations of the study also need to be noted. First and foremost, this study focuses more on the ease of use and operability of evaluation system. In the future research, the main task is to establish a reasonable and effective measurement system for college's performance based on rigorous and scientific methods, meanwhile, both practicality and reliable theoretical basis for measurement system must be taken into account, which will provide a solid foundation for the research; Secondly, the future research should carry out a longitudinal study for a single alternative proposal, as well as monitor and explore the measurement indicators that affect the performance of colleges using time series data, so that colleges can implement their own improvements in a targeted manner.

Author contribution statement

Lei-Yi Peng: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Le-Ying Pan: Analyzed and interpreted the data; Wrote the paper.

Jia Lu: Contributed reagents, materials, analysis tools or data.

Yan-Man Long: Conceived and designed the experiments.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data included in article/supp. material/referenced in article.

Declaration of interest's statement

The authors declare no competing interests.

References

[1] A. Djangone, O. El-Gayar, An empirical study of the impact of knowledge acquisition, sharing and utilization on organizational performance of higher education institutions (HEIs), and the moderating role of organizational culture, Issues Inf. Syst. 22 (4) (2021) 258–268.

[2] J.G. Vargas Hernández, S.T. Cano Ibarra, Evaluating higher education institutions through agency and resources-capabilities theories: a model for measuring the perceived quality of service, Indepen. J. Manag. Prod. 7 (4) (2016) 1126–1153.

[3] F.J. Mhwise, The impact of total quality management critical success factors in higher education institutions performance in Iraq, Univ. Thi-Qar J. 16 (1) (2021) 1–14.

[4] M. Mohammadi, Dimensions of teacher performance evaluation by students in higher education, Shanlax Int. J. Educ. 9 (2) (2021) 18–25.

[5] M.L. Bougnol, J.H. Dulá, Validating DEA as a ranking tool: an application of DEA to assess performance in higher education, Ann. Oper. Res. 145 (1) (2006) 339–365.

[6] S. Nazari-shirkouhi, S. Mousakhani, M. Tavakoli, M. Reza Dalvand, J. Šaparauskas, J. Antuchevičienė, Importance-performance analysis based balanced scorecard for performance evaluation in higher education institutions: an integrated fuzzy approach, J. Bus. Econ. Manag. 21 (3) (2020) 647–678.

[7] F.F. Al-Hosaini, S. Sofian, A review of balanced scorecard framework in higher education institution (HEIs), Int. Rev. Manag. Market. 5 (1) (2015) 26–35.

[8] T.L. Saaty, D. Ergu, When is a decision-making method trustworthy? Criteria for evaluating multi-criteria decision-making methods, Int. J. Inf. Technol. Decis. Making 14 (2015) 1171–1187.

[9] K. Govindan, S. Rajendran, J. Sarkis, P. Murugesan, Multi criteria decision making approaches for green supplier evaluation and selection: a literature review, J. Clean. Prod. 98 (2015) 66–83.

[10] H. Wang, L. Chen, K. Chen, M.Q. Xue, Q. Liang, Multi-index comprehensive evaluation method and selection of weight coefficient, J. Guangdong Univ. Pharm. 23 (5) (2007) 583–589.

[11] B.M. Beamon, Measuring supply chain performance, Int. J. Oper. Prod. Manag. 19 (3) (1999) 275–292.

[12] M. Franco-Santos, M. Kennerley, P. Micheli, V. Martinez, S. Mason, B. Marr, D. Grey, A. Neely, Towards a definition of a business performance measurement system, Int. J. Oper. Prod. Manag. 27 (8) (2007) 784–801.

[13] P. Leahy, J.M. Salazar, The moral dimension in Chilean higher education's expansion, High Educ. 74 (1) (2017) 147–161.

[14] H.Y. Wu, J.K. Chen, I.S. Chen, H.H. Zhuo, Ranking universities based on performance evaluation by a hybrid MCDM model, Measurement 45 (5) (2012) 856–880.

[15] S.H. Zolfani, A.S. Ghadikolaei, Performance evaluation of private universities based on balanced scorecard: empirical study based on Iran, J. Bus. Econ. Manag. 14 (4) (2013) 696–714.

[16] E. Bayraktar, E. Tatoglu, S. Zaim, Measuring the relative efficiency of quality management practices in Turkish public and private universities, J. Oper. Res. Soc. 64 (2013) 1810–1830.

[17] R.R. Duh, K.T. Chen, R.C. Lin, L.C. Kuo, Do internal controls improve operating efficiency of universities? Ann. Oper. Res. 221 (1) (2014) 173–195.

[18] J. Nazarko, J. Šaparauskas, Application of DEA method in efficiency evaluation of public higher education institutions, Technol. Econ. Dev. Econ. 20 (1) (2014) 25–44.

[19] L. Ding, Y. Zeng, Evaluation of Chinese higher education by TOPSIS and IEW-The case of 68 universities belonging to the Ministry of Education in China, China Econ. Rev. 36 (2015) 341–358.

[20] L. Dai, J. Li, Study on the quality of private university education based on analytic hierarchy process and fuzzy comprehensive evaluation method 1, J. Intell. Fuzzy Syst. 31 (4) (2016) 2241–2247.

[21] G. Yang, H. Fukuyama, Y. Song, Measuring the inefficiency of Chinese research universities based on a two-stage network DEA model, J. Inf. 12 (1) (2018) 10–30.

[22] X.H. Xu, L. Ye, Y. Pei, L. Zhao, J.J. Wang, Research on the comprehensive evaluation of the higher education system based on FCE and ARMA models, Complexity (2022), 3142579.

[23] J. Texeira-Quiros, M.R. Justino, M.G. Antunes, P.R. Mucharreira, A.T. Nunes, Effects of innovation, total quality management, and internationalization on organizational performance of higher education institutions, Front. Psychol. 13 (2022), 869638.

[24] F. Ecer, D. Pamucar, A. Mardani, M. Alrasheedi, Assessment of renewable energy resources using new interval rough number extension of the level based weight assessment and combinative distance based assessment, Renew. Energy 170 (2021) 1156–1177.

[25] S.J. Teixeira, J.J. Ferreira, P. Wanke, A.J.J. Moreira, Evaluation model of competitive and innovative tourism practices based on information entropy and alternative criteria weight, Tourism Econ.: Bus. Fin. Tour. Recreat. 27 (2021) 23–44.

[26] P. Wang, P. Meng, J.Y. Zhai, Z.Q. Zhu, A hybrid method using experiment design and grey relational analysis for multiple criteria decision making problems, Knowl. Base Syst. 53 (2013) 100–107.

[27] G.W. Wei, Grey relational analysis model for dynamic hybrid multiple attribute decision making, Knowl.-Based Syst. 24 (5) (2011) 672–679.

[28] N. Senthilkumar, T. Tamizharasan, V. Anandakrishnan, Experimental investigation and performance analysis of cemented carbide inserts of different geometries using Taguchi based grey relational analysis, Measurement 58 (2014) 520–536.

[29] T. Bretag, R. Harper, M. Burton, C. Ellis, P. Newton, P. Rozenberg, Contract cheating: a survey of Australian university students, Stud. High Educ. 44 (11) (2019) 1837–1856.

[30] S. Tannock, No grades in higher education now! Revisiting the place of graded assessment in the reimagination of the public university, Stud. High Educ. 42 (8) (2017) 1345–1357.

[31] R. Wolf, D. Zaher, R. Benjamin, Methodological challenges in international comparative post-secondary assessment program: lessons learned and the road ahead, Stud. High Educ. 40 (3) (2015) 471–481.

[32] B. Margaret, A. Rola, Can a rubric do more than be transparent? Invitation as a new metaphor for assessment criteria, Stud. High Educ. 46 (2) (2021) 359–368.

[33] S.E. Hikichi, E.G. Salgado, L.A. Baijo, Forecasting number of ISO 14001 certifications in the Americas using ARMA models, J. Clean. Prod. 147 (2017) 242–253.

[34] J. Watrobski, J. Jankowski, P. Ziemba, A. Karczmarczyk, M. Ziolo, Generalised framework for multi-criteria method selection, Omega 86 (2019) 107–124.

[35] A.R. Mishra, P. Rani, K. Pandey, Fermatean fuzzy CRITIC-EDAS approach for the selection of sustainable third-party reverse logistics providers using improved generalized score function, J. Ambient Intell. Hum. Comput. (2021) 1–17.

[36] A. Ishikawa, M. Amagasa, T. Shiga, G. Tomizawa, R. Tatsuta, H. Mieno, The max-min Delphi method and fuzzy Delphi method via fuzzy integration, Fuzzy Set Syst. 55 (3) (1993) 241–253.

[37] C.E. Shannon, A mathematical theory of communication, The Bell Syst. Tech. J. 27 (1948) 379–423.

[38] I. Mukhametzyanov, Specific character of objective methods for determining weights of criteria in MCDM problems: entropy, CRITIC and SD, Appl. Manag. Eng. 4 (2021) 76–105.

[39] S. Monghasemi, M.R. Nikoo, F.M.A. Khaksar, J. Adamowski, A novel multi criteria decision making model for optimizing time–cost–quality trade-off problems in construction projects, Expert Syst. Appl. 42 (2015) 3089–3104.
[40] T. Parveen, H.D. Arora, M. Alam, Intuitionistic Fuzzy Shannon Entropy Weight Based Multi-Criteria Decision Model with TOPSIS to Analyze Security Risks and Select Online Transaction Method, in: Advances in Computing and Intelligent Systems, Springer, 2020, pp. 1–17.
[41] S. Lu, Y. Shang, W. Li, Assessment of the Tarim River basin water resources sustainable utilization based on entropy weight set pair theory, Water Supply 19 (3) (2019) 908–917.
[42] X. Liang, W. Liang, L. Zhang, X. Guo, Risk assessment for long distance gas pipe-lines in coal mine gobs based on structure entropy weight method and multi-step backward cloud transformation algorithm based on sampling with replacement, J. Clean. Prod. 227 (2019) 218–228.
[43] M. Sacasqui, J. Luyo, A. Delgado, A Unified Index for Power Quality Assessment in Distributed Generation Systems Using Grey Clustering and Entropy Weight, in: 2018 IEEE ANDESCON, IEEE, 2018.
[44] O. Mohsen, N. Fereshteh, An extended VIKOR method based on entropy measure for the failure modes risk assessment: a case study of the geothermal power plant (GPP), Saf. Sci. 92 (2017) 160–172.
[45] R.K. Mavi, M. Goh, N.K. Mavi, Supplier selection with Shannon entropy and fuzzy TOPSIS in the context of supply chain risk management, Proc.-Soc. Behav. Sci. 235 (2016) 216–225.
[46] N.A. Maidin, S.M. Sapuan, M.M. Taha, M.Z.M. Yusoff, Material selection of natural fibre using a grey relational analysis (GRA) approach, Bioresources 17 (1) (2022) 109–131.
[47] G.W. Wei, GRA method for multiple attribute decision making with incomplete weight information in intuitionistic fuzzy setting, Knowl. Base Syst. 23 (3) (2010) 243–247.
[48] J.L. Deng, Spread of grey relational space, J. Grey Syst. 7 (1995) 96–100.
[49] Y.C. Hu, Tolerance rough sets for pattern classification using multiple grey single-layer perceptrons, Neruocomputing 179 (29) (2016) 144–151.
[50] A.N. Patil, N.G. Pai Bhale, N. Raikar, M. Prabhakaran, Car selection using hybrid fuzzy AHP and grey relation analysis approach, Int. J. Perform. Eng. 13 (5) (2017) 569–576.
[51] M. Ikram, R. Sroufe, E. Rehman, S.Z.A. Shah, A. Mahmoudi, Do quality, environmental, and social (qes) certifications improve international trade? a comparative grey relation analysis of developing vs. Developed Countries, Phys. A Stat. Mech. Its Appl 545 (2020) 1–34.
[52] S.A. Javed, S. Liu, Bidirectional absolute GRA/GIA model for uncertain systems: application in project management, IEEE Access 7 (2019) 60885–60896.
[53] A. Suvvari, P. Goyari, Financial performance assessment using Grey relational analysis (GRA), Grey Syst. Theor. Appl. 9 (2019) 502–516.
[54] E. Aydemir, Y. Sahin, Evaluation of healthcare service quality factors using grey relational analysis in a dialysis center, Grey Syst. Theor. Appl. 9 (2019) 432–448.
[55] H. Malekpoor, K. Chalvatzis, N. Mishra, M.K. Mehlawat, D. Zafirakis, M. Song, Integrated grey relational analysis and multi objective grey linear programming for sustainable electricity generation planning, Ann. Oper. Res. 269 (1–2) (2018) 475–503.
[56] B. Zhang, Z. Shen, R. Ma, J. Yan, M. Yu, Grey relational analysis and fuzzy synthetic discrimination of antioxidant components in peach fruit, Arch. Biol. Sci. 70 (3) (2018) 449–458.
[57] N. Uckun, T. Kose, N. Girginer, Combined use of data development analysis and grey relation analysis in the efficiency analysis of public hospitals in the metropolitan cities of Turkey, Int. J. Soc. Sci. Econ. Res. 2 (9) (2017) 4570–4590.
[58] K.E. Newcomer, Using performance measurement to improve programs, N. Dir. Eval. 75 (1997) 5–14.
[59] X.D. Peng, H. Garg, Intuitionistic fuzzy soft decision making method based on CoCoSo and CRITIC for CCN cache placement strategy selection, Artif. Intell. Rev. 55 (2022) 1567–1604.
[60] K. Lisiecka, E. Czyż-Gwiazda, Performance Measurement Models - Comparative Review, in: 57 Th EOQ Congress: Quality Renaissance - Co-creating a Viable Future, Tallinn, Estonia, 2013.
[61] M. Tsinidou, V. Gerogiannis, P. Fitsilis, Evaluation of the factors that determine quality in higher education: an empirical study, Qual. Assur. Educ. 18 (3) (2010) 227–244.
[62] S.A.H. Syed Hassan, S.C. Tan, K.M. Yusof, MCDM for Engineering Education: Literature Review and Research Issues, in: M. Auer, K.S. Kim (Eds.), Engineering Education for a Smart Society. GEDC WEEF 2016 2016. Advances in Intelligent Systems and Computing, Springer, Cham, 2018, p. 627.
[63] G.J. Klir, T.A. Folger, Fuzzy Sets, Uncertainty, and Information, Prentice Hall, USA, 1988.
[64] Y.F. Kuo, P.C. Chen, Constructing performance appraisal indicators for mobility of the service industries using Fuzzy Delphi method, Expert Syst. Appl. 35 (2008) 1930–1939.
[65] X.Y. Chen, J.X. Wang, X.F. Zhang, A study on the key core resources and evaluation indicators for the development of off-site restaurants in sightseeing hotels, Hospit. Sightseeing 14 (4) (2017) 223–249.
[66] X.R. Tan, Y.G. Li, Using grey relational analysis to analyze the medical data, Kybernetes 33 (2) (2004) 355–362.
[67] L.C. Hsu, S.L. Ou, Y.C. Ou, A comprehensive performance evaluation and ranking methodology under a sustainable development perspective, J. Bus. Econ. Manag. 16 (1) (2015) 74–92.
[68] D. Luo, Analysis Method of Grey Decision-Making Problem, Yellow River Water Conservancy Press, Zhenzhou, 2005.
[69] C.Y. Fu, J.S. Zheng, J.M. Zhao, Application of Grey Relational Analysis for corrosion failure of oil tubes, Corrosion Sci. 43 (5) (2001) 881–889.
[70] P.L. Williams, C. Webb, The Delphi technique: a methodological discussion, J. Adv. Nurs. 19 (1) (1994) 180–186.
[71] G. Skulmoski, F. Hartman, J. Krahn, The Delphi Method for graduate research, J. Inf. Technol. Educ. 6 (1) (2007) 1–21.
[72] C.E. Huang, Y. Yuan, C.F. Huang, Differences between public and private institutions of Taiwan's HTVE system in determinants of competitiveness, US China Educ. Rev. 5 (7) (2008) 1–12.