International Journal of Organizational Innovation

A STUDY ON THE APPLICATION OF THE CCR MODEL FOR THE CLASSIFICATION OF UNIVERSITY TEACHERS' PERFORMANCE

Hsiang-Ching Lai*,
Dept. of Industrial Management, I- Shou University, Kaohsiung 84001, Taiwan (ROC)
*Corresponding Author: xiagshein@nuk.edu.tw

Chiao-Pin Bao
Dept. of Industrial Management, I- Shou University, Kaohsiung 84001, Taiwan (ROC)
cpbaao@cloud.isu.edu.tw

Chun-I Chen
Dept. of Industrial Management, I- Shou University, Kaohsiung 84001, Taiwan (ROC)
EddyChen@isu.edu.tw

Abstract

This paper discusses how to effectively classify the performance of university teachers in order to obtain the information of personnel traits, make the individual get feedback on professional development, and make up for the deficiency of the rank of personnel. This study explores two classification methods: Method 1 “Z-score classification” (the size of the Z-score in research, teaching and service determines the classification criteria); Method 2 “CCR model classification” (classification of various scores). We simulated the performance scores of 134 teachers in a university in three categories: research, teaching and service and used two classification methods. Result 1: all the research scores obtained by "Z-score classification" and "CCR model classification" were significantly different in research, teaching and service categories through ANOVA; all teaching scores obtained by the two methods were significantly different in research, teaching and service categories through ANOVA; the service scores obtained by the two methods were significantly different in research, teaching and service categories through ANOVA. Result 2: according to the t-test, there was no significant difference in the research scores in the research category, teaching scores in the teaching category and service scores in the service category based on the two classification methods, indicating that both methods were effective. Result 3: According to the F test (f_0=2.1769>1.0) of the total variation of Method 1 and Method 2, the total variation of Method 1 is much greater than that of Method 2, indicating that Method 2 is better than Method 1.

Keywords: Z-score, CCR model, classification

Introduction

Professional evaluation of university teachers is considered to be a key factor related to the professional development of teachers, personnel management, and quality of university education (Mesak and Jauch, 1991; Hyle, 1999). There has been a high level of global attention to the development of teacher performance evaluation (Mitchell and Leachman, 2015), and the development of fair and effective models has been a challenge. According to the literature gap analysis, with a lack of effective, composite, practical and flexible evaluation models for teachers, their professional competencies cannot be distinguished, thus affecting their development in teaching, research and service activities (Askar, 2019). Teacher evaluation is a very specialized, complex and specialized practice. For a long time, the process has been neglected. The re-evaluation tends to be self-evaluation by teachers, and there is a lack of an effective method (Liang and Ouyang, 2017).

When discussing the differences in the performance of university teachers in various professions, Shifflett and Patterson (1995) believed that it is a challenge to highlight the balance between teaching, research, academia, consulting services and administration in teachers' professional duties. There is neither a clear definition of the nature of research, teaching and service, nor a consensus discussion and appropriate evaluation (Price and Cotton, 2006; Gentry, 2013). Due to various academic training and professional development in universities, as well as the independent personality traits and professional development types of teachers, it is difficult to obtain consistent results in teacher evaluation (Braskamp and Qry,1994; Weistroffer, Spinelli, Canavos, and Fuhs, 2001).

Performance evaluation is a sensitive subject for teachers and schools, requiring fair, clear standards, various forms of evaluation tools, multiple assessments, frequent and useful feedback, the establishment of teacher evaluation systems, and the provision of personal development decisions for teachers (Sayavedra, 2014). The teacher evaluation system is designed to identify teacher development and growth (Marzano, 2012; Dee and Wyckoff, 2015). The teacher evaluation system is a key factor in improving teachers' professional performance (Tuytens and Devos, 2011; Delvaux, Vanhoof, Tuyten, Vekeman Devos, and Petegem, 2013). According to different categories of teacher performance evaluation, each evaluation should be effectively described and explained to form evaluation criteria for individual characteristics, and an evaluation system and standard for feature selection should be established to promote appropriate and reasonable evaluation (Liang and Ouyang, 2017). Poor evaluation interactions (psychological trust between evaluators and evaluatees) result from inadequate fairness and accurate and just evaluation systems (Santiago, Roseveare, Amelsvoort, Manzi, & Matthews, 2009).

The inaccuracy and inconsistency in the teacher performance evaluation model (Darling-Hammond, Amrein-Beardsley, Haertel, and Rothstein, 2012), makes it impossible to accurately measure teachers' ability and effectively distinguish teachers' efficiency (Marzano, 2012; Bill and Melinda Gates Foundation, 2011; Toch and Rothman, 2008; U.S. Department of Education, 2009; Weisberg, Sexton, Mulhern, and Keeling, 2009).

Literature Review

There are many theoretical bases from the evaluation system model design, evaluation standards and teachers' professional performance level, but there are many empirical limitations. Relevant literature is summarized as follows:

Evaluation of Data Mining Technique in Predicting Teachers' Performance

Pal (2013) used data mining tools such as Naive Bayes and ID3 to evaluate 13 variables for teachers, processed some indefinable performance data, and distinguished teachers' performance with the five-point Likert scale. The accuracy rate of Naive Bayes is the highest, but it is only 80.35%.

Asanbe, Osoisan, and William (2016) used Artificial Neural Network and decision tree to calculate teachers' work experience, level, qualification, contract status, professional qualification and other attributes to predict their performance. However, different data mining techniques have different classification accuracy, so follow-up empirical research is still needed. Shanmugara-Jeshwari and Lawrance (2017) used decision tree classification technology to analyze teachers' performance, established the attributes of observation data with 38 kinds of data and 10 kinds of ability data of teachers, and classified teachers into five groups of performance. However, the establishment of the system underwent missing value removal and classification, and 60%-70% of the time in the data analysis process was spent on “data preprocessing”. Therefore, without good data, the subsequent analysis will be highly biased.

Application of Quantitative and Qualitative Research of the Cloud Model to Teacher Performance Evaluation

Quantitative data refer to statistical data with an objective structure supported by specific data. Meredith, (Steward, and Lewis 2011; Fitchett, and Heafner 2017). Qualitative data refer to non-measurement data that describes the subject as orientation (Raoul, Bergstrom and Mann, 2006; Carlos and Oliverira, 2012). It is undeniable that the quantitative and qualitative performance evaluation of teachers should consider cross-criteria, potential overlap and double counting. Chang and Wang (2015) used the cloud model to deal with quantitative and qualitative problems and uncertain evaluation data, but the connotation of quantitative and qualitative data was not clearly defined in the research process. The challenge of this model is the research on similarity measurement and multi-dimensional performance, and different evaluation criteria are required for problems in different fields (Yang, Wang, and Liu, 2018).

Application of the Combination of Hierarchical Analysis and Delphi with Expert Optimization Weight Assignment to Teacher Performance Evaluation

For any type of university personnel, the proper balance between research, teaching and service has not been clearly established (Grant and Fogarty, 1998; Costa and Oliveira, 2012), and it is difficult to assign specific values to these activities and find appropriate indicators to accurately measure performance (Adler and Harzing, 2009). Optimizing and setting the elastic weight of the project standard can reflect the value tradeoff between the criteria, but whether evaluation of the weight through weighting, optimization and correction procedures is appropriate is the common problem in research (Costa and Oliveira, 2012; Keeney, See and Winterfeldt, 2006). According to Weistroffer et al. (2001), various activities of teachers in teaching, research and service have been specifically set, but these items will be revised to adapt to the department and environmental objectives. Therefore, the evaluator should have the right to revise the weight to deal with special situations or system models that cannot be predicted or suggested. Chang and Wang (2015) determined the weight of pointer items by combining the hierarchical analysis and Delphi method, optimized and calculated the average weight, and finally carried out the consistency test. These methods used in weight assignment are not only unable to obtain accurate empirical results, but also require the subjective will of people to be evaluated.

Breakpoint Regression is Used to Differentiate Teacher Performance

To respect the functions and characteristics of universities, the Ministry of Education allows universities to set the methods and criteria for teacher performance evaluation, and the evaluation of teachers by universities is mainly based on the improvement of teaching, research and service (Wang and Chen, 2005). Weisberg et al. (2009) pointed out that the failure of a TNTP program in the United States was due to the fact that less than 1% of the teachers were rated as "unsatisfied" level, which was obviously inconsistent with the results of the education supervisor's inspection that 57% of the teachers were ineffective. Holtzapple (2003) used discontinuous breakpoint regression design to divide 395 teachers' performance from 2001 to 2002 and 393 teachers' performance from 2002 to 2003 into "dissatisfied", "general", "proficient" and "excellent" levels according to students' learning achievement. The research shows that it is only more sensitive in identifying "dissatisfied" teachers. Teachers get high scores for their annual performance, and the performance grades with low differentiation cannot effectively distinguish teachers' performance (Koedel, Li, Springer,and Tan, 2018; Kraft, Gilmour, and Allison, 2017; Weisberg et al., 2009). Many teachers' performance is rated as "satisfied" and "unsatisfied", and the two performance levels alone cannot distinguish teachers' performance. Even when the evaluation results are divided into more than three levels, many teachers are still rated at the highest level and thus have poor discrimination (Weisberg et al., 2009; Fang and Fang, 2017). The literature list related to university teacher evaluation is summarized as shown in Table 1.

The data envelopment analysis (DEA) developed by Charnes, Cooper, and Rhodes (1978) is a nonparametric approach for evaluating the performance of a set of decision making units (DMUs) which use multiple inputs to generate multiple outputs.

According to Wang, Luo and Lan (2011), the traditional DEA-CCR model enables DMUs to assess their maximum efficiency score with the most favorable weights. But, it has some drawbacks; for instance, the efficiencies of different DMUs gained by different sets of weights may fail to be compared and ranked on the same basis. Moreover, there are always more than one DMU to be measured as efficient because of the flexibility in the selection of weights; it would cause the situation that all DMUs cannot be fully discriminated.

As indicated by Emrouznejad & Yang (2018), DEA theory and its applications fall within the fields of management science and operations research. Besides, agriculture, banking, supply chain, transportation and public policies are the top 5 fields of DEA application. A great number of journal articles on related topics were published in 2015 and 2016.

Nurmatov, Lopez and Millan (2021) discovered the most commonly used DEA methods (i.e. CCR and BBC models) account for 20.41% of the related research in 2017 and 2018. Also, the literature relevant to other DEA models (virtual frontier DEA model; CCR-DEA model; meta-frontier DEA; hybrid DEA model; robust DEA model; super efficiency DEA model; stochastic DEA model; BBC-DEA model) accounts for 30.61%. To conclude, only little literature focused on exploring the application of the traditional CCR-DEA model in classification. However, the combination of the DEA method and deep learning in classification research is still confronted with several constraints. Thus, more data are required to verify its efficiency.

The innovation of this study lies in the adoption of the CCR model to minimize the situations that the weights of weak items often equal 0. This study highlights the classification of the CCR model can be applied to a wide range of fields. This study uses the example of teachers' performance to verify the efficiency

Research Method

The purpose of performance evaluation is mainly to promote teachers' professional development. It is a sensitive topic for teachers and universities It needs fair and clear standards, various forms of evaluation tools to conduct multiple evaluations, regular and useful feedback, and provide suggestions for teachers' personal development (Sayavedra, 2014). The following two innovative performance classification methods are proposed in order to enable university teachers to obtain an effective classification in professional

Table 1. Brief description of teacher evaluation literature

Category	Researcher / year	Advantages
Evaluation of Data Mining Technique in predicting teachers' performance	Pal (2013) Asanbe et al. (2016) Shanmugarajeshwari and Lawrance (2017)	1. Data mining technology attempts to establish an effective model to evaluate teachers' performance based on a large number of different factors and overcome the limitations of traditional methods. 2. Data mining is defined as the process of exploring and analyzing large amounts of data in an automatic or semi-automatic manner, usually discovering meaningful patterns and rules.
		Disadvantages
		1. Different data mining techniques have different classification accuracy, so follow-up empirical research is still needed. 2. The establishment of the system undergoes missing value removal and classification, and 60%-70% of the time in the data analysis process is spent on "data preprocessing". Without good data, the subsequent analysis will be highly biased.
Application of quantitative and qualitative research of the cloud model to teacher performance evaluation	Arreola (2007) Meredith et al. (2011) Costa and Oliveira (2012) Chang and Wang (2015) Yang et al. (2018)	Advantages
		The cloud model can fully express the fuzzy and random language, and the allocation of qualitative and quantitative data in the retrieval language is represented by value.
		Disadvantages
		1. The cross-criteria, potential overlap and double calculation should be considered in quantitative and qualitative performance. 2. Different fields require different evaluation criteria. It is difficult to assign specific values to these activities and find appropriate indicators to accurately measure performance.
Valuation of teacher perform-	Grant and Fogarty (1998), Weistroffer et	Advantages
		The weight is a scaling constant that can

ance by weight assignment	al. (2001) Adler and Harzing (2009) Keeney et al. (2006) Costa and Oliveira (2012)	be used to aggregate value scores according to different criteria. Therefore, its evaluation requires a trade-off of value.
		Disadvantages
		1. The use of weights is not completely accurate. It is difficult to assign specific values to these activities and find appropriate indicators to accurately measure performance 2. The appropriate and inappropriate evaluation of the weight through weighting, optimization and correction procedures is the common problem in research. 3. These methods used in weight assignment are not only unable to obtain accurate empirical results, but also require the subjective will of people to be evaluated.
Distinguishing teachers' performance levels	Holtzapple (2004) Koedel et al. (2018) Kraft et al. (2017) Fang and Fang (2017) Weisberg et al. (2009)	Advantages
		It leads to the development of more rigorous and informative evaluation systems.
		Disadvantages
		1. Teachers' performance differences cannot be distinguished effectively. 2. The consistency of evaluators' views on university teachers' effectiveness with their actual performance grades is concerned.

Note. Compiled by this study

performance evaluation and appropriate professional development and suggestions.

Method 1 "Z-score Classification":

After standardization, the score of each evaluated teacher in research, teaching and service is taken to determine the classification standard according to the value of the Z-score obtained.

The person with the highest Z-score is assigned to this group.

$$\text{Eq.} > \quad max\left\{\frac{x_{ij}-\bar{x}_j}{S_j}\right\}, j = A, B, C \qquad \text{...(1)}$$

x_{ij} indicates the score of the ith teacher in the jth category.

$\bar{x}_j$ is the average of all teachers in

Category j.

*S*j is the standard deviation of each teacher's score in the jth category.

A, B and C respectively represents the three categories of university teachers: research, teaching and service.

Method 2: "CCR model classification".

It adopts the CCR model of Data Envelopment Analysis (DEA), as shown in Eq. 2. The weight of "0" is used as the criterion of grouping. First, the highest score of the research category minus the original scores of all research teachers. The highest score for teaching items minus all teacher scores in the teaching category. The same goes for the service category. The new variable data obtained are analyzed by the CCR model, and the items with the weight of 0 are regarded as the weak items of the evaluatees, otherwise it is the advantage item:

1. In the CCR model, the weight may be 0 because there is no criterion for the constraint. When the weight of the teacher's evaluated item is "0", then he or she is classified as that item.
2. If there are more than two evaluated items with the weight of "0", the score with the smallest variation of the item will be used as the classification standard.
3. If there is no weight of "0" in the three items, it is assumed that it is the "i" item, and Equation (3) below will be used as the criterion for selection:

The linear programming of the DEA-CCR model is as follows (Charnes, Cooper, and Rhodes, 1978):

Eq.>

$$\text{Max } \theta_k = \sum_{r=1}^{s} u_{rk} y_{rk}$$

$$\text{s.t.} \sum_{i=1}^{m} v_{ik} \quad x_{ik} = 1$$

$$\sum_{i=1}^{s} u_{rj} y_{rj} - \sum_{r=1}^{m} v_{ij} x_{ij} \leq 0, k = 1,2 \ldots, n$$

$$i = 1,2, \ldots, s, r = 1,2 \ldots, m \quad \ldots\ldots\ldots\ldots (2)$$

$\theta_k = Efficiency\ value\ of\ DMU_K$, K=1,2,…,n.

u_{rk} is *the output weight of* DMU_k

v_{ik} is the input weight of DMU_k

y_{rk} is the output r variable of DMU_k

x_{ik} is the iutput i variable of DMU_k

y_{rj} is the output r variable of DMU_j

x_{ij} is the intput i variable of DMU_j

Eq.>

$$min\left\{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j\ )^2\right\}, j = A, B, C, n = 134 \ldots\ldots (3)$$

x_{ij} is the score of the ith teacher in the jth category; $\bar{x}_j$ is the average score of all teachers in in the jth category; A, B, C represents the three categories of university teachers respectively: research, teaching and service.

Research Process

Stage 1 Research Process

Z-score classification and CCR model classification were used to perform single-factor ANOVA and ex-post t-test for teachers' scores, as shown in Figure 1.

Stage 2 Research Process

The research group obtained by Z-score classification and the research group obtained by CCR model classification, the teaching group obtained by Z-score classification and the teaching group obtained by CCR model classification, and the service group obtained by Z-score classification and the service group obtained by CCR model classification were all t-tested, as shown in Figure 2.

Stage 3 Research Process

F-test was used to test the total variation of Z-score classification and CCR model classification, as shown in Figure 3.

Case Analysis

Phase 1

Method 1: "Z-score classification" The Z-score was obtained by simulating the scores of 134 teachers in research, teaching and service categories in a university. Those with the highest Z-scores were divided into this group. The grouping results are shown in Table 2.

After "Z-score classification" grouping, the research scores of teachers in the research, teaching and service categories were analyzed by ANOVA and ex-post t-test:

The research scores of teachers in research, teaching and service categories were significantly different in group efficiency by one-way ANOVA (P*=0.001<.05). The analysis is shown in Table 3.

The result of the t-test for teachers in research and teaching categories is P*<0.001<.05, and there is a significant difference between the two groups. The analysis is shown in Table 4.

The result of the t-test for teachers in research and service categories is P*<0.001<.05, and there is a significant difference between the two groups. The analysis is shown in Table 5.

The result of the t-test for teachers in teaching and service categories is P=0.39>.05, and there is no significant difference between the two groups. The analysis is shown in Table 6.

After "Z-score classification" grouping, the teaching scores of teachers in the research, teaching and service categories were analyzed by ANOVA and ex-post t-test. The following is an analysis table (omitted):

The result of one-way ANOVA for the three groups is P*=3.36E-09<.05, which shows that there is a significant difference among the three groups.

The result of the t-test for teachers in research and teaching categories is P*=3.87E-09<.05, and there is a significant difference between the two groups.

The result of the t-test for teachers in research and service categories is P=0.22>.05, and there is no significant difference between the two groups.

Figure 1. Stage 1 process

The result of the t-test for teachers in teaching and service categories is P*=1.83E-06<.05, and there is a significant difference between the two groups.

After "Z-score classification" grouping, the service scores of teachers in the research, teaching and service categories were analyzed by ANOVA and ex-post t-test:

The result of one-way ANOVA for the three groups is P*=5.81E-11<.05, which shows that there is a significant difference among the three groups.

The result of the t-test for teachers in research and teaching categories is P=0.4565□.05, and there is no significant difference between the two groups.

Table 2. Grouping results of "Z-score classification" for 134 teachers in research, teaching and service majors in a university

No.	Group	No.	Group	No.	Group	No.	Group	No.	Group
1	A	31	A	61	C	91	A	121	A
2	A	32	B	62	B	92	A	122	B
3	A	33	A	63	B	93	A	123	A
4	A	34	A	64	C	94	A	124	A
5	B	35	B	65	B	95	B	125	B
6	C	36	C	66	A	96	A	126	A
7	C	37	C	67	A	97	C	127	A
8	B	38	B	68	B	98	B	128	B
9	B	39	B	69	B	99	B	129	B
10	A	40	C	70	B	100	C	130	C
11	B	41	B	71	B	101	C	131	B
12	B	42	C	72	C	102	C	132	C
13	B	43	B	73	B	103	C	133	B
14	C	44	C	74	A	104	C	134	C
15	C	45	C	75	C	105	C		
16	A	46	A	76	A	106	C		
17	A	47	B	77	A	107	A		
18	C	48	C	78	B	108	C		
19	A	39	A	79	C	109	C		
20	A	50	B	80	A	110	A		
21	B	51	C	81	A	111	C		
22	B	52	C	82	A	112	B		
23	B	53	C	83	C	113	B		
24	B	54	C	84	C	114	C		
25	C	55	B	85	B	115	A		
26	A	56	C	86	C	116	C		
27	A	57	A	87	C	117	B		
28	B	58	B	88	B	118	B		
29	B	59	B	89	B	119	A		
30	C	60	B	90	B	120	C		

Note. A, B and C represent the three categories of research, teaching and service for university teachers.

The result of the t-test for teachers in research and service categories is P*=1.12E-06 <.05, and there is a significant difference between the two groups.

The result of the t-test for teachers in teaching and service categories is P*=3.92E-08<.05, and there is a significant difference between the two groups.

Figure 2. Stage 2 research process

Figure 3. Stage 3 research process

Method 2: "CCR model classification" adopts the CCR- DEA model, as shown in Eq. (2). The weight of "0" is used as the criterion of grouping. The weight may be 0 because here is no criterion for the constraint. The highest score of the item minus the original score of each teacher from research, teaching and service fields. The items with the weight of 0 are regarded as the weak items of the evaluatees, otherwise it is the advantage item. For example, the original score 822 of the 126th teacher in the research category is the highest, the original score 458 of the 72nd teacher in the teaching category is the highest, and the score 698 of the 72nd teacher in the service category is the highest. We use 822 to subtract the research score of each evaluated teacher, 458 to subtract the teaching score of each evaluated teacher, and 698 to subtract the service score of all teachers. If there are more than two items with the weight of "0", the one with the least variation in the score (such as Eq. 3) will be used as the classification standard. The grouping results of each participant are shown in Table 7.

Grouping Analysis of CCR Model Classification

After "CCR model classification", the researching scores of teachers in the research, teaching and service categories were analyzed by one-way ANOVA and ex-post t-test.

Table 3. ANOVA for the research scores of research, teaching, and service categories by Z-score classification

Variable source	*SS*	*df*	*MS*	*F*	*P*	Critical value
Between-group	1621150	2	810574.95	47.25	<0.001	3.07
Within-group	2247268	131	17154.72			
Total	3868418	133				

Note. SS= Sum of squares; df= degree of freedom; MS= Mean square; F= MS_B/MS_W; P^* <.05

Table 4. T-test of research scores for teachers in research and teaching categories.

Item	Number	mean	Variance	*df*	*P*	*t*
Research score of the research category	38	306.66	56674.88	85	<0.001	6.95
Research score of the teaching category	49	67.98	1050.73			

Note. df= degree of freedom; P* <.05

Table 5. T-test of research scores for teachers in research and service categories.

Item	Number	mean	Variance	*df*	*P*	*t*
Research score of the research category	39	309.56	1043.89	82	<0.001	6.80
Research score of the service category	45	66.22	1968.86			

Note. df= degree of freedom; P* <.05

Table 6. T-test of research scores for teachers in teaching and service categories.

Item	Number	mean	Variance	*df*	*P*	*t*
Research score of the teaching category	50	68.52	1043.89	93	0.39	0.29
Research score of the service category	45	66.22	1968.86			

Note. df= degree of freedom; P >.05

Table 7. Grouping results of the "CCR model" for 134 teachers in a university

No.	Group	No.	Group	No.	Group	No.	Group	No.	Group
1	B	31	A	61	B	91	B	121	B
2	A	32	B	62	B	92	A	122	B
3	B	33	B	63	B	93	B	123	B
4	A	34	A	64	C	94	C	124	B
5	B	35	B	65	B	95	B	125	B
6	C	36	A	66	A	96	C	126	B
7	C	37	C	67	B	97	C	127	B
8	B	38	B	68	B	98	B	128	B
9	B	39	B	69	B	99	B	129	B
10	A	40	B	70	B	100	B	130	B
11	B	41	B	71	B	101	B	131	B
12	C	42	C	72	B	102	C	132	C
13	B	43	B	73	B	103	C	133	B
14	B	44	B	74	B	104	C	134	C
15	C	45	C	75	C	105	C		
16	A	46	B	76	B	106	B		
17	B	47	B	77	B	107	A		
18	C	48	B	78	B	108	C		
19	B	39	B	79	A	109	C		
20	C	50	B	80	C	110	A		
21	B	51	C	81	A	111	C		
22	A	52	A	82	A	112	B		
23	B	53	C	83	C	113	B		
24	B	54	B	84	C	114	C		
25	B	55	B	85	B	115	B		
26	C	56	C	86	C	116	B		
27	C	57	A	87	C	117	B		
28	B	58	B	88	B	118	B		
29	B	59	B	89	B	119	B		
30	B	60	B	90	B	120	C		

Note. A, B, C represent the three categories of university teachers from research, teaching and service fields.

Table 8. ANOVA for the research scores of research, teaching, and service categories by CCR model classification

Variable source	*SS*	*df*	*MS*	*F*	*P*	Critical value
Between-group	358085.83	2	179042.92	6.68	0.02	3.07
Within-group	3510331.90	131	26796.43			
Total	3868418	133				

Note. SS= Sum of squares; df= degree of freedom; MS= Mean square; F= MS_B/MS_W; P^* <.05

Table 9. T-test of research scores for teachers in research and service categories

Item	Number	mean	Variance	*df*	*P*	*t*
Research score of the research category	17	258.30	26650.10	50	<0.001	5.25
Research score of the service category	35	81.43	6586.25			

Note. df= degree of freedom; P* <.05

Table 10. T-test of research scores for teachers in research and teaching categories

Item	Number	Mean	Variance	*df*	*P*	*t*
Research score of the research category	17	258.29	26650.10	97	0.01	2.48
Research score of the teaching category	82	137.05	35308.61			

Note. df= degree of freedom; P* <.05

Table 11. T-test of research scores for teachers in teaching and service categories

Item	Number	mean	Variance	*df*	*P*	*t*
Research score of the teaching category	82	137.05	35308.61	115	0.05	1.68
Research score of the service category	35	81.43	6586.25			

Note. df= degree of freedom; P□.05

The research scores of teachers in research, teaching and service categories were significantly different in group efficiency by one-way ANOVA (P*=0.02<.05). The analysis is shown in Table 8.

The result of the t-test for teachers in research and teaching categories is P*=0.01<.05, and there is a significant difference between the two groups. The analysis is shown in Table 9.

The result of the t-test for teachers in research and service categories is P*<0.001<.05, and there is a significant difference between the two groups. The analysis is shown in Table 10.

The result of the t-test for teachers in teaching and service categories is P=0.47□.05, and there is no significant difference between the two groups. The analysis is shown in Table 11.

After "CCR model classification" grouping, the teaching scores of teachers in the research, teaching and service categories were analyzed by ANOVA and ex-post t-test. The following is an analysis table (omitted):

The result of one-way ANOVA for the three groups is P*=2.54E-08<.05, which shows that there is a significant difference among the three groups.

The result of the t-test for teachers in research and teaching categories is P*=0.0073<.05, and there is a significant difference between the two groups.

The result of the t-test for teachers in research and service categories is P*= 0.0201<.05*, and there is a significant difference between the two groups.

The result of the t-test for teachers in teaching and service categories is P*= 1.60E-09 <.05, and there is a significant difference between the two groups.

After "CCR model classification" grouping, the service scores of teachers in the research, teaching and service categories were analyzed by ANOVA and ex-post t-test:

The result of one-way ANOVA for the three groups is P*=0.0301<.05, which shows that there is a significant difference among the three groups.

The result of the t-test for teachers in research and teaching categories is P=0.3523>.05, and there is no significant difference between the two groups.

The result of the t-test for teachers in research and service categories is P*=0.0335<.05, and there is a significant difference between the two groups.

The result of the t-test for teachers in teaching and service categories is P*= 0.0058<.05, and there is a significant difference between the two groups.

The results of "CCR model classification" in ANOVA and t-test are better than that of "Z-score classification". CCR model classification has a better discriminative tendency than Z-score classification.

Phase 2:

Difference test after grouping of "Z-score classification" and "CCR model classification":

(1) The t-test was used for the research scores of the research category by the two methods.

After the t-test of the research group, P=0.2096>.05, the two groups showed no significant difference.

(2) The t-test was used for the teaching scores of the teaching category by the two methods.

After the t-test of the teaching group, P=0.06>.05, the two groups showed no significant difference.

(3) The t-test was used for the service scores of the service category by the two methods.

After the t-test of the service group, P=0.051>.05, the two groups showed no significant difference.

(4) The t-test results for the grouping of "Z-score classification" and "CCR model classification" showed no significant difference, indicating that the grouping effects of the two classifications were similar.

Phase 3:

The total variation of "Z-score classification" was 3,028,173, which was much larger than that of "CCR model classification", which was 1,391,033. Therefore, "CCR model classification" has a better discriminative tendency than "Z-score classification"

$$= \frac{3028173}{1391033} = 2.1769 > F_{0.00} \ {}_{(188,188)} \ 1.0$$

The total variation of the two methods was significantly different, so the CCR method was better than the standardized classification method.

Suggestions

The results of this study can be used as a reference for policies related to education administration. Teachers in the research category should be encouraged to invest in research by reducing teaching hours to increase their capacity and output. Teachers in the teaching category should increase the teaching hours so that their achievements can be brought into full play in the learning performance of students and stimulate excellent development in teaching. Teachers in the service category should be encouraged to take up educational administrators and supervisors, so as to promote the administrative development of educational organizations by accumulating administrative experience. Compared with the classification method of performance comparison and ranking, this type of performance evaluation can indeed give teachers feedback on appropriate professional development, and universities can also get specific suggestions on how to make full use of their talents and effective human resource management.

Derrington and Kirk (2017) probed the concept of teacher-as-learner and adopted four elements for job-embedded learning (teacher as learner-focused, community-focused, assessment-focused, and knowledge-focused) to help teachers with their professional development.

Darling-Hammond, and Hyler (2017) reviewed 35 methods and concluded the following 7 features for effective professional development: (1) Is content focused; (2) Incorporates active learning; (3) Supports collaboration; (4) Uses models of effective practice; (5) Provides coaching and expert support; (6) Offers feedback and reflection; and (7) Is of sustained duration.

For teachers' professional development, the CCR classification proposed by this research can provide effective suggestions for teachers' professional classification. Whether it is employed in summative evaluation or formative evaluation, this taxonomy shows the theoretical spirit for teachers' performance evaluation by connecting teachers' professional development. Besides, it can also provide feedback to universities and teachers to enable cooperation, gain supports from experts, put into practice, and serve as a reference.

Conclusions

Guba and Lincoln (1989) and Ding (2006) emphasized that the importance of evaluation is a process of dialogue and negotiation based on the respect for the personality and uniqueness of people to be evaluated. In the past literature, the emphasis on weight assignment was not only unable to obtain accurate empirical data, but also not in line with the subjective will of the people to be evaluated.

On the basis of more respect for the personality and uniqueness of people to be evaluated, the characteristics of the CCR model lie in the use and emphasis of their strengths. The purpose of this study is to develop the classification tool of teachers' performance, so that teachers can obtain appropriate professional performance classification, and specific suggestions and effective feedback for future professional development can be provided. With the respect and attention of the evaluated people's subjectivity, the classification of teachers' professional performance is taken as the result of performance evaluation, instead of the previous emphasis on distinguishing teachers' performance levels or performance ranking, so that the quality of teacher evaluation is diversified in value. Respect for differences is the basic trait of subjective evaluation (Liang, 2007; Ding, 2006).

In this study, two different approaches were proposed to classify teachers' expertise:

The results of "Z-score classification" showed significant differences in research scores, teaching scores, and service scores among research, teaching, and service categories by ANOVA. After the t-test, the research scores for teaching and service categories, the teaching scores for research and service categories, the service scores for research and teaching categories are not significant, but the other groups are significant. The results of "CCR model classification" showed significant differences in research scores, teaching scores and service scores for research, teaching and service categories by one-way ANOVA. After the t-test, the research scores for teaching and service categories, the service scores for research and teaching categories are not significant, but the other groups are significant. According to ANOVA and t-test results, "CCR model classification" has a better discriminative tendency than "Z-score classification".

The research, teaching and service groups were classified by the two classification methods. According to the t-test, there is no significant difference, which indicates that the two methods are effective in teacher classification.

The total variation of Z-score classification is greater than that of the CCR efficiency evaluation model classification. The total variation ratio of the two classification methods is
$F_0 = 2.1769 > F_{0.00\ (133,133)} \cong 1.0$
, so Method 2 is better than Method 1.

Both of the two classification methods can achieve grouping benefits. Based on the simulation data situation in this study, it is found that "CCR Model Classification" is better than "Z-score Classification", so it is appropriate for the university to adopt "CCR Model Classification" to evaluate teachers' performance.

It is often considered that the weight of "0" is a major defect of the CCR efficiency evaluation model. In this study, the weight of "0" was used as an effective classification tool, which is a feature of the study.

In this study, the CCR efficiency model was used to evaluate teachers' performance. When the weight of one variable was "0", teachers were classified into this category. However, if there are two variable weights of "0", the teacher performance scores cannot be classified, and the "standardized scores" will be implemented according to the mean of each class and the standard deviation. The average, standard deviation of various categories, and perhaps the large dispersion of the original data may lead to an inaccurate evaluation of Z-score classification. To fix this shortcoming, it is suggested to further explore in the future.

References

Adler N. J., & Harzing, A. W. (2009). When knowledge wins: Transcending the sense and nonsense of academic rankings. *Academy of Management Learning and Education*, *8*, 72-95.

Asanbe, M. O., Osofisan, A. O., & William, W. F. (2016). Teachers' performance evaluation in higher educational institution using data mining technique. *International Journal of Applied Information Systems, 10*(7), 10-15.

Askar, M. (2019). Faculty target-based engagement assessment statistical model for enhancing performance and education quality. *IAFOR Journal of Education, 7*(2), 27-49.

Bill and Melinda Gates Foundation. (2011). Learning about teaching: Initial findings from the Measures of Effective Teaching project.

Braskamp, L. A., & Qry, J.C. (1994). Assessing faculty work. San Francisco: Jossey-Bass.

Chang, T. C., & Wang, H. (2015). A multi criteria group decision-making model for teacher evaluation in higher education based on tree. *Eurasia Journal of Mathematics, Science and Technology Education, 12*(5), 1243-1262.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429–444.

Costa, B.E., Carlos. A. & Oliveira, M. D. (2012). *A multicriteria decision analysis model for faculty evaluation. Omega, 40*(4), 424-436.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluation teacher evaluation. *Phi Delta Kappan, 93*(6), 8-15.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). Effective teacher professional development. Learning Policy Institute, 1-76.

Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management, 34*(2), 267-297.

Delvaux, E., Vanhoof, J., Tuytens, M., Vekeman, E., Devos, G., & Van Petegem, P. (2013). How may teacher evaluation have an impact on professional development? A multilevel analysis. *Teaching and Teacher Education, 36*, 1-11.

Derrington, M. L., & Kirk, J. (2017). Linking job-embedded professional development and mandated teacher evaluation: Teacher as learner. *Professional Development in Education, 43*(4), 630-644.

Ding, Y. K. (2006). Preliminary discussion on the planning and implementation of the teacher advancement system: on the evaluation of teacher professional development. *Secondary Education, 57*(5), 60-73.

Emrouznejad, A., & Yang, G. L. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-economic planning sciences, 61*, 4-8.

Fang, C. Y., & Fang, D. L. (2017). “Teacher evaluation 2.0”: Learning for Experiences of the United States of America and implications for Taiwan. *Journal of Educational Research and Development, 13*(1), 31-62.

Fitchett, P. G., & Heafner, T. L. (2017). Quantitative research and large-scale secondary analysis in social studies. The Wiley handbook of social studies research, 68.

Gentry, R. (2013). Ensuring that professors who enhance the university earn tenure and promotion. Online Submission.

Grant, J., & Fogarty, T. (1998). Faculty evaluation as a social dilemma: A game theoretic approach. *Accounting Education, 7*(3), 225-248.

Guba, E. G., & Lincoln, Y. S. (1989). Fourth generation evaluation. Sage.

Holtzapple, E. (2003). Criterion-related validity evidence for a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education, 17*(3), 207-219.

Hyle, A. E. (1999). Faculty evaluation: A prickly pair. *Higher Education, 38*, 351-371.

Keeney, R. L., See, K. E., Winterfeldt, D. (2006). Evaluating academic programs: With applications to U.S. graduate decision science programs. *Operations Research, 54*, 813-828.

Koedel, C., Li, J., Springer, M. G., & Tan, L. (2018). Teacher performance ratings and professional improvement. *Journal of Research on Educational Effectiveness, 12*(1), 90-115.

Liang, H. J. (2007). Differentiated teacher evaluation. Shanghai: East China Normal University.

Liang, Z., & Ouyang, W. P. (2017). A model of teacher evaluation system based improved fuzzy clustering algorithm. *Revista de la Facultad de Ingenieria U.C.V., 32*(7), 461-466.

Marzano, R. J. (2012). Teacher evaluation. *Educational leadership, 70*(3) ,14-19.

Meredith, J.R., Steward, M.D., & Lewis, B.R. (2011). Knowledge dissemination in operations management: Published perceptions versus academic reality. *OMEGA-The International Journal of Management Science, 39*, 435-446.

Mesak, H. I., & Jauch, L. R. (1991). Faculty performance evaluation: Modeling to improve personnel decisions. *Decision Sciences, 22*(5), 1142-1157.

Mitchell, M., & Leachman, M. (2015). Years of cuts threaten to put the college out of reach for more students. Center on Budget and Policy Priorities, 13.

Nurmatov, R., Lopez, X. L. F., & Millan, P. P. C. (2021). Tourism, hospitality, and DEA: Where do we come from and where do we go? *International Journal of Hospitality Management, 95*, 102883.

Pal, S. (2013). Evaluation of teacher's performance: A data mining approach. *International Journal of Computing Science and Mobile Computing, 2*(12), 359-639.

Price, J., & Cotton, S. R. (2006). Teaching, research and service: Expectations of assistant professors. *The American Sociologist, 379*(1), 5-21.

Raoul, A., Bergstrom, O., & Mann, B. (2006). Developing a comprehensive faculty evaluation system: A guide to designing, building, and operating large-scale faculty evaluation systems. Jossey-Bass.

Santiago, P., Roseveare, D., Van Amelsvoort, G., Manzi, J. & Matthews, P. (2009). Teacher evaluation in Portugal OECD Review, Citeseer, 3-89.

Sayavedra, M. (2014). Teacher evaluation. *ORTESOL Journal, 31*, 1-9.

Shanmugarajeshwari, V., Lawrance, R., (2017). An analysis of teachers' performance using classification techniques.

Shifflett, B., & Patterson, P. (1995). Succeeding in the tenure and promotion process. *Physical Educator, 52*(3), 160-169.

Toch, T., & Rothman, R. (2008). Rush to judgment: Teacher evaluation in public education. Education Sector Reports. Education Sector.

Tuytens, M., & Devos, G. (2010). The influence of school leadership on teachers' perception of teacher evaluation policy. *Education Studies, 36*(5), 521-536.

U.S. Department of Education. (2009). Race to the top executive summary. accessed, available at from www2.ed.gov/programs/racetothetop/executive-summary pdf. Sun, J., Wu, J., & Guo, D. (2013). Performance ranking of units considering ideal and anti-ideal DMU with common weights. Applied Mathematical Modelling, 37(9), 6301-6310.November, 1-15.

Wang, C. S., & Chen, P. T. (2005). A teacher evaluation system for university faculties: A case study of a National University in Taiwan. *Journal of Technology Management, 10*(3), 121-152.

Wang, Y. M., Luo, Y., & Lan, Y. X. (2011). Common weights for fully ranking decision making units by regression analysis. *Expert Systems with Applications, 38*(8), 9122-9128.

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. The New Teacher Project.

Weistroffer, H. R., Spinelli, M. A., Canavos, C. A., & Fuhs, F. P. (2001). A merit pay allocation model for college faculty based on performance quality and quantity. *Economics of Education Review, Nurmatov, R., Lopez, X. L. F., & Millan, P. P. C. (2021). Tourism, hospitality, and DEA: Where do we come from and where do we go?. International Journal of Hospitality Management, 95, 102883.20*(1), 41-49.

Yang, C. et al. (2018). Review and prospect of normal cloud models. *Chinese Journal of Computers, 41*(3), 24-744.