Expert Systems with Applications 38 (2011) 14542–14549

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A learning performance evaluation with benchmarking concept for English writing courses

Bernard Montoneri [a,*], Chia-Chi Lee [b], Tyrone T. Lin [c], Shio-Ling Huang [c]

[a] Department of English Language, Literature and Linguistics, Providence University, Taiwan 200, Chung Chi Rd., Taichung 43301, Taiwan, ROC
[b] Department of Accounting Information, National Taipei College of Business, Taiwan 321, Sec. 1, Jinan Rd., Zhongzheng District, Taipei 100, Taiwan, ROC
[c] Department of International Business, National Dong Hwa University, Taiwan 1, Sec. 2, Da Hsueh Rd., Shou-Feng, Hualien 974, Taiwan, ROC

ARTICLE INFO

Keywords:
Data enveloped analysis (DEA)
English writing
Facet reference sets
Learning performance benchmarking

ABSTRACT

This paper adopts data envelopment analysis (DEA), a robust and reliable evaluation method widely applied in various fields to explore the key indicators contributing to the learning performance of English freshmen writing courses in a university of Taiwan from the academic year 2004 to 2006. The results of DEA model applied in learning performance change our original viewpoint and reveal that some decision-making units (DMUs) with higher actual values of inputs and outputs have lower efficiency because the relative efficiency of each DMU is measured by their distance to the efficiency frontier. DMUs may refer to different facet reference sets according to their actual values located in lower or higher ranges. In the managerial strategy of educational field, the paper can encourage inefficient DMUs to always compare themselves with efficient DMUs in their range and make improvement little by little. The results of DEA model can also give clear indicators and the percentage of which input and output items to improve. The paper also demonstrates that the benchmarking characteristics of the DEA model can automatically segment all the DMUs into different levels based on the indicators fed into the performance evaluation mechanism. The efficient DMUs on the frontier curve can be considered as the boundaries of the classification which are systematically defined by the DEA model according to the statistic distribution.

1. Introduction

According to the Ministry of Education in Taiwan (2009a), the number of students in higher education institutions (HEIs) has increased by about 23% from 915,921 to 1,123,755 from 1999 to 2009. During the past decade, higher education has become accessible to more people with the encouragement from the Ministry of Education. As a result, the number of colleges and universities has increased rapidly. However, the government budget has not increased as much as the number of academic institutions. Many of these institutions struggle to obtain public financial support and need to find additional financial support.

Moreover, even if higher education is now universal, Taiwan has the lowest birth rate in the world: 0.83 according to the Ministry of Interior (2010). The birth rate has dropped dramatically during the past 4 years certainly due to the global economic downturn. If this trend continues, the number of students in Taiwan will inevitably decrease in the coming years. The percentage of children under 15 years of age has already dropped from 20.80% in 2001 to 16.95% in 2008 (Ministry of Education, 2009b) and the number of elementary school children has dropped from 1,927,179 during 1999–2000 to 1,677,303 during 2008–2009 (Ministry of Education 2001, 2009a).

In order to distribute limited resources more fairly and close less competitive institutions, the Higher Education Evaluation and Accreditation Council of Taiwan has initiated a system of performance evaluation based on self-evaluation reports and accompanied by peer reviews or site visits (Ministry of Education, 2009a). Higher education institutions need to maintain a certain level of quality and become as efficient and attractive as possible in order to avoid low student enrollment, high graduate unemployment, credential inflation, and even closure. The performance evaluation by the Ministry of Education is now a universal practice undertaken every four years in Taiwan. Every academic institution also performs self-evaluations concerning teachers' research results and student's learning performance.

This paper adopts data envelopment analysis (DEA), a robust and reliable evaluation method widely applied in various fields to explore the key indicators contributing to the learning performance of English freshmen writing courses in a university of Taiwan from the academic year 2004 to 2006. Taiwan export-based economy makes English an indispensable communication tool and

* Corresponding author. Tel.: +886 426328001x12021; fax: +886 4 26334712.
E-mail addresses: bmon@pu.edu.tw (B. Montoneri), cclee.vera@msa.hinet.net (C.-C. Lee), tjlin@mail.ndhu.edu.tw (T.T. Lin), slhuang@mail.ndhu.edu.tw (S.-L. Huang).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2011.05.029

a valuable skill for the students who expect to enter the job market. The paper aims at not only calculating the quantitative efficiency of DMUs (evaluated units), but also indicating the room for improvement in what aspect for the inefficient DMUs. Through the benchmarking characteristics of DEA, the inefficient DMUs can have more objective indication and endeavor to make progress step by step.

According to McGowan and Graham (2009), the four factors contributing most to improved teaching are active/practical learning (real-world experiences and in-class discussions), teacher/student interactions (knowing students personally), clear expectations/learning outcomes, and faculty preparation. By comparison in our paper, the following four factors are classified into two inputs and two outputs: the preparation of teaching contents and teaching skills for the inputs and fair grading and students' learning performance for the outputs. Active/practical learning may correspond to our teaching skills and faculty preparation to our preparation of teaching contents. However, contrary to McGowan and Graham (2009), the paper shows that fair grading and students' learning performance are the major factors contributing to students' learning performance and teachers' improved teaching.

Dickinson (1990) made a study in an American public school by using Mitzel's, 1982 criterion for classifying the rating questions. Students were asked about instructors' knowledge, presentation, overall evaluation, presentation of clear objectives, and about students' estimate of amount learned. His paper determined the relationship between learning and the students' ratings of the instruction by using Pearson correlation coefficients. The findings pointed out the problems with using pupils' ratings of teachers especially if the ratings were to be used as part of a teacher's evaluation for merit pay, promotion, or tenure if amount learned was used as the criterion of effective teaching: according to Chacko (1983), Arreola (1986) and Dickinson (1990), and students generally gave their teachers high marks even in the face of low learning.

Marsh and Roche (1997) showed that student evaluation could result in improved teaching effectiveness and that end-of-term feedback is more effective than midterm feedback. The validity of student evaluation is demonstrated by validation of student ratings, student achievement, teacher self-evaluation, and observation by trained evaluators. But not all scholars acknowledge the value of evaluating faculty performance. Ryan, Anderson, and Birchler (1980) showed that 38% of professors admitted to making their courses easier in response to student evaluation. Boland and Sims (1988) criticized the process as subjective, inconsistent, punitive, and sporadic. They showed that evaluation criteria were inconsistently written and that performance expectations were inexplicit or poorly communicated. Haskell (1997) analyzed the impact of student evaluation of faculty performance on academic freedom and argued that the pressure to respond to students' needs could cause the problems regarding retention, promotion, tenure, and salary increases.

Students' ratings of teachers are part of the performance evaluation. Even though these ratings are not considered by some scholars to be very objective, however, they provide useful information and give an opportunity to students to express their feelings and what they think about their teachers and their courses. It should only be used as a reference, but with great care and fairness. Moreover, the questions asked to the students can be improved in order to make the results more accurate and objective. When the ratings are unfair and do not reflect the reality, it will make teachers feel bitter and want to give up. When the ratings are accurate and fair, it can help teachers better understand their students and their needs, leading to improved teaching methods.

Data envelopment analysis (DEA) has been used by various studies to analyze students' learning performance. Ahn, Arnold, Charnes, and Cooper (1989) have assessed the efficiency of US universities during 1981–1985 and Glass, Mckillop, and O'Roruke (1998) the efficiency of UK universities during 1989–1992. Other studies have measured efficiency at the departmental level: Madden, Savage, and Kemp (1997) measured the performance of economics departments in Australian universities; Johnes and Johnes (1993) the performance of economics departments in the UK during 1984–1988; Colbert, Levary, and Shaner (2000) the performance of MBA programs in the US. Fu and Huang (2009) conducted a survey of college graduates in 2003 and used an output-oriented BCC type of DEA model to provide information for students and school administrators.

Lin (2009) developed an evaluation approach for measuring and ranking the efficiency of tutors in some higher education institutions (HEIs) in Taiwan. As many of these institutions have established committees to evaluate the performance of tutors and reward the outstanding tutors each academic year, Lin (2009) proposed to use an IDEA (imprecise data envelopment analysis) model based on the BCC model to help determining the final ranking of outstanding DMUs, that is, the evaluated tutors.

In the field of Teaching English as a Second Language (TESL), Barcelos and Kalaja (2003) demonstrated that teachers' and students' beliefs about second language acquisition are experiential, dynamic, socially constructed, and changeable. Hsu (2010) proposed a web-based interactive speaking improvement system for English as a Second Language learner by using fuzzy matching. According to this study, the system could effectively help students speak English more correctly and it could be adapted for improving the speaking ability of learners of other foreign languages than English. Clinton and Kohlmeyer (2005) have investigated the effect of group quizzes on accounting students' performance and motivation to learn. They found out that students in the group-condition had significantly different affective reactions than those in the non-group condition. Moreover, the overall performance of the instructor was rated higher for the group quizzes sections and the students expressed greater motivation to learn and increased enthusiasm even if they did not evidence any significantly different performance results.

The remainder of the paper is organized as follows: the methodology and choice of key indicators explain the DEA method and the important input and output items discussed in the paper. The following section presents the obtained numerical results and recommendations. The final paragraph draws the conclusions and implications.

2. Methodology and choice of inputs and outputs

Performance evaluation has been widely applied in various industries and fields. The objective of evaluation aims at enhancing management performance, changing strategy, or increasing productivity. Enterprises and institutions use assessment tools or methods to clarify the performance of people, machines, equipment resources, and operation procedures. Therefore, managers can better allocate limited resources and make optimal strategies and decisions in time. For the organizations possessing human resources as their major assets, employees' working performance and learning performance are also fatal in a fierce competitive market. Education institutions even consider training high quality students both in professional knowledge and in daily life behavior as one of their most important missions. The students going to university have different background and characteristics every year. It is quite difficult to predict their learning performance, especially because English writing courses are taught by many teachers providing different teaching efforts. The paper aims at identifying the main input indicators having a significant impact on the output indicators. The input and output indicators describing this type of learning performance should better be quantitative so as to be compared with different evaluated units. The DEA method is particularly suitable and reliable for this type of study.

2.1. DEA model, Charnes–Cooper–Rhodes model and Banker–Charnes–Cooper model

Data envelopment analysis (DEA) is a quantitative method which can handle both multiple inputs and multiple outputs to measure the performance of decision-making units (DMUs), that is, the evaluated units. DEA includes the function and concept of benchmarking. The inefficient DMUs can refer to the outstanding DMUs on their range and are not always compared with the ones with the highest range of actual values. The DEA method can indicate the relative efficiency of each DMU within a sample (Samoilenko & Osei-Bryson, 2008). Charnes, Cooper, and Rhodes (1978) converted the concept of multiple inputs and multiple outputs into single virtual input and output by linear combination. Their method, called the "Charnes–Cooper–Rhodes (CCR) model" or "CCR model" estimated the efficiency frontier by the ratio of two linear combinations and measured the relative efficiency of each DMU. According to Lin, Lee, and Chiu (2009) and Lee (2009), the efficiency value of the CCR model corresponds to the overall technical efficiency of an evaluated unit. If the efficiency value equals 1, the evaluated unit is efficient (optimal performance) and in constant returns to scale (CRTS); if the efficiency value is less than 1, the evaluated unit needs some improvement. Banker, Charnes, and Cooper (1984) expanded the concept and application scope in both Farrell and CCR models by supposing that the inefficiency of a DMU might not have allocative efficiency, proper scale, and technical efficiency. Therefore, they changed the CCR model's CRTS hypothesis to variable returns to scale (VRS) and divided technical efficiency into pure technical efficiency and scale efficiency. This method is called the "Banker–Charnes–Cooper (BCC) model" or "BCC model".

2.2. Definition of data source and DMU

The paper adopts DEA to perform the efficiency evaluation of freshmen students in a university of Taiwan from the academic year 2004 to 2006. They follow the same training program of English writing for one semester. A total of 50 classes are selected as the decision-making units (DMUs), that is, the evaluated units. They are named from D1 to D50. The learning performance of the DMUs is then interpreted by analyzing input items and output items.

2.3. Selection of input and output items

The specification of inputs and outputs is a crucial first step in DEA. According to students' ratings, each question is given a score from 1 to 5. 1 means that students are very unsatisfied; 2 means that students are unsatisfied; 3 means that students are neither satisfied nor unsatisfied; 4 means that students are satisfied; 5 means that students are very satisfied. Four items for the evaluation model are chosen as follows:

Input dimension:

I1 The preparation of teaching contents: it can reflect the degree of teachers' professional knowledge for the preparation of teaching materials. A high score signifies that students are highly satisfied with teachers' preparation of teaching contents.

I2 Teaching skills: students acknowledge their teacher's ability to make the course of English writing clear and interesting and have no difficulty to assimilate the contents. A high score signifies that students are highly satisfied with teachers' teaching skills.

Output dimension:

O1 Fair grading: it shows whether teachers' grading is fair according to students. A higher score denotes that students consider teachers' grading to be fairer.

O2 Students' learning performance: it indicates students' knowledge acquisition after a semester of English writing training. A higher score denotes that students are more satisfied with what and how much they learned over the past semester.

The average actual values (students' ratings of teachers) of I1, I2, O1, and O2 are 3.97, 3.82, 3.83, and 3.99, respectively. All the acquired data are treated and then interpreted in the following sections.

2.4. Correlation analysis of input and output items

The paper uses the Pearson correlation coefficient test to verify the correlation of inputs and outputs so as to understand whether the principle of Isotonicity is satisfied. A higher Pearson correlation coefficient means that the relationship between two variables is closer. A Pearson correlation coefficient of 0.8 or above indicates a very high correlation; a value between 0.6 and 0.8 indicates a high correlation; a value between 0.2 and 0.4 indicates a low correlation; a value inferior to 0.2 indicates an extremely low correlation or no correlation. Table 1 lists the correlation coefficients between input (I) and output (O) items: I1 represents the preparation of teaching contents and I2 teaching skills; O1 represents fair grading and O2 students' learning performance. The correlation coefficients among these 4 items are all above 0.9 with a significant level at 1%. Therefore, the principle of Isotonicity is satisfied. In addition, the variance inflation factor (VIF) check is undertaken to measure the impact of collinearity between the two inputs and between the two outputs. The result shows that the VIF values all satisfy the norm (VIF < 10); that is, the VIF value inferior to 10 represents that there is no highly collinearity problem between the inputs or between the outputs, and that the high correlation problem does not exist in our paper. Hence, the inputs and outputs chosen in the paper can not replace one another; each item is representative of the evaluation of learning performance.

3. Results and recommendations

The results of numerical analysis may provide useful suggestions for both teachers and students. These results can help students' learning efforts and teachers' teaching efforts to reach their optimal performance under limited resources. In the paper, the software Frontier Analyst is used to analyze the learning performance of 50 DMUs, named from D1 to D50. The calculi are undertaken respectively by using an input oriented and an output oriented DEA model to evaluate the use efficiency of relative resources. An input orientation evaluates the minimum input effort needed to maintain the current output performance, while an

Table 1
Correlation coefficients between input and output items.

Inputs outputs	I1 (preparation of teaching contents)	I2 (teaching skill)
O1 (Fair grading)	0.961***	0.939***
O2 (Learning performance)	0.936***	0.908***

*** Denotes significant levels at 1%.

output orientation evaluates the maximum output performance needed under the current input resources.

In our paper about learning performance in higher educational institutions, minimizing input items in order to obtain an efficiency value equal to 1 can mislead educators because:

1. In order to obtain a high overall technical efficiency, some teachers choose to reduce their teaching efforts and ignore the improvement of their teaching skills. A study by Ryan et al. (1980) showed that 38% of professors admitted making their courses easier in response to student evaluation.
2. It is possible for a teacher to obtain a very high overall technical efficiency by reducing the preparation of teaching contents and by giving higher grades to students. This kind of attitude will probably lead to the decline of students' professional knowledge.
3. If hard-working teachers know that it is possible to obtain a higher evaluation by reducing the preparation of teaching contents, it can discourage them from making tremendous efforts to better prepare their courses and improve their teaching skills.

Therefore, the input oriented model is not suitable for our paper because it is contrary to the objective of education. If used, this input oriented model must follow the criteria which can avoid the problems mentioned above. Teaching resources and teachers' energy are limited. Hence, the output oriented model is more suitable because it emphasizes on how much the insufficiency of the output performance is under the current input resources without additional input efforts.

We have undertaken the calculi of both the input and output oriented BCC models (assuming returns to scale are variable). The results show that the changes in efficiency have the same tendency. The DMUs with an efficiency value of 1 are the same for both input and output orientations; but for the DMUs with an efficiency value inferior to 1, there are slight gaps. The same phenomenon appears in the result of other indicators probably because the goal-settings and limitations of mathematical model for both input and output orientations are different. Therefore, only the output oriented empirical results and some important indicators are chosen and listed in Table 2. In addition, Table 2 also lists the CCR score, that is, the overall technical efficiency of 50 DMUs calculated under the CCR model of DEA in order to highlight the differences of efficiency values between the CCR model and BCC model. The BCC model improves the CCR model and assumes that the DMUs are scale variable, which is more objective, reasonable, and realistic. Therefore, the results of the other parameters obtained under the output oriented BCC model are showed in Table 2. The DMUs are ranked by descending order of "CCR" and BCC model's "Refs" columns. The column "Refs" indicates the number of times the other DMUs are referring to it.

3.1. Efficiency analysis of learning performance

The DMUs with an efficiency value equal to 1 are efficient and can constitute "facet reference sets" which form efficiency frontier curves. These efficient DMUs, on the frontier curves, become referring standards for other inefficient DMUs. The efficiency value of each DMU is calculated by the distance of their locations to the efficiency frontier curves. The DMUs in the top 50% of the CCR scores are arranged in the left part of Table 2 with the average efficiency of 0.982; in the right part of Table 2 are the other DMUs with the average efficiency of 0.941. We observe that five DMUs have an overall technical efficiency value of 1, including D37, D41, D22, D6, and D49. They represent 10% of all the evaluated units; they have the best performance and are ranked from 1 to 5. These 5 DMUs do not need any improvement in the input items or in the output items because they have reached their optimal state, meaning that both teachers and students feel at ease and are motivated to work. This state requires a good atmosphere in the class and good teaching preparation and skills from the part of the teachers. At the same time, students have to show their efforts during the training and to accept criticism. They should not be afraid of making mistakes and accept to be corrected. Sometimes, students do not like to be too much corrected and teachers have to know how much amount of criticism they can handle. This attitude belongs to "teaching skills". If students are over criticized, they will probably loose their motivation.

According to the BCC model, DMUs are variable returns to scale (VRS) and the overall technical efficiency can be divided into pure technical efficiency and scale efficiency, named "BCC" and "Scale" in Table 2 respectively. Under managerial discretion, pure technical efficiency helps to know whether the production cost has been minimized and whether the output efficiency has reached its optimal performance. That is, whether the input resources are perfectly used or lay idle. In our paper, the pure technical efficiency means whether the preparation of teaching contents and teaching skills (input items) are sufficiently presented during the courses.

The scale efficiency is also called allocative efficiency or price efficiency, which refers to the ability of finding an optimal ratio for a configuration of fixed inputs and outputs. That is, the products should reach their economic scale in order to cost down the production. The input resources and output performance are weighed at the same time. In general, the input cost is higher than the value on the minimum cost line; therefore, the scale efficiency, calculated by the minimum cost over the real cost is always inferior or equal to 1. In our paper, the scale efficiency inferior to 1 means most teachers prepare more teaching contents, but students cannot assimilate or appreciate 100% of the teaching contents or even are not aware of them.

The 5 DMUs with the overall technical efficiency value of 1 have the pure technical efficiency value and scale efficiency value of 1, which indicates that they have reached their optimal state. Among the other DMUs in Table 2. 6 do not attain the optimal overall technical efficiency but obtain the pure technical efficiency value of 1 (including D25, D29, D45, D16, D12, and D10). These six DMUs' scale efficiency values are all smaller than 1, indicating that the inefficiency of these DMUs mainly stems from the scale factor. Furthermore, there are 39 inefficient DMUs like D3 whose pure technical efficiency values are smaller than 1. Among which, 6 DMUs (D3, D28, D11, D21, D30, and D26) have the pure technical efficiency values greater than the scale efficiency values, indicating that the inefficiency of these DMUs is mainly due to the fact that students are unable to assimilate the entire contents of the courses. As to the other inefficient 33 DMUs, their pure technical efficiency values are smaller than scale efficiency values. It means that some students think that their teacher has not enough professional knowledge and experience to teach a course of English writing.

The "RTS" column in Table 2 indicates DMUs' state of returns to scale. The constant returns to scale (CRTS) represents that the DMUs' inputs and outputs reach a state of optimal configuration without the need of any adjustments. The increasing returns to scale (IRTS) represents that DMUs can increase the input resources in order to increase the output performance. The decreasing returns to scale (DRTS) represents that DMUs should reduce the input resources and adjust the output configuration. Table 2 shows that among 39 pure technical inefficient DMUs, 26 inefficient DMUs like D3 are in the stage of increasing returns to scale (IRTS). The other inefficient 13 DMUs like D11 are in the stage of decreasing returns to scale (DRTS) and recommended to reduce the quantity of teaching contents and improve teaching skills.

Table 2
Output oriented evaluation indicators of 50 DMUs.

ID	Ranking	CCR	BCC	Scale	RTS	Refs	Peers	Room for improvement (%)				ID	Ranking	CCR	BCC	Scale	RTS	Refs	Peers	Room for improvement (%)			
								O1	O2	I1	I2									O1	O2	I1	I2
D37	1	1.000	1.000	1.000	CRTS	25	0	0	0	0	0	D30	26	0.963	0.986	0.977	IRTS	0	3	1.5	1.5	0	−4
D41	2	1.000	1.000	1.000	CRTS	21	0	0	0	0	0	D17	26	0.963	0.975	0.988	IRTS	0	2	2.6	3.3	0	−0.7
D22	3	1.000	1.000	1.000	CRTS	13	0	0	0	0	0	D15	28	0.961	0.968	0.992	DRTS	0	4	3.3	3.3	0	0
D6	4	1.000	1.000	1.000	CRTS	9	0	0	0	0	0	D44	29	0.953	0.958	0.995	IRTS	0	3	4.4	4.4	0	−0.2
D49	5	1.000	1.000	1.000	CRTS	1	0	0	0	0	0	D47	30	0.953	0.955	0.998	DRTS	0	4	4.7	4.7	0	0
D25	6	0.999	1.000	0.999	CRTS	2	0	0	0	0	0	D28	31	0.951	0.994	0.956	IRTS	0	2	0.6	3.3	−2.9	0
D29	7	0.997	1.000	0.997	CRTS	2	0	0	0	0	0	D48	31	0.951	0.975	0.976	IRTS	0	2	2.6	13.7	−0.1	0
D45	8	0.995	1.000	0.995	CRTS	24	0	0	0	0	0	D9	33	0.949	0.962	0.987	IRTS	0	3	6.6	4	0	0
D38	9	0.987	0.988	0.999	IRTS	0	4	1.2	1.2	0	0	D19	34	0.948	0.956	0.991	IRTS	0	4	4.6	4.6	0	0
D16	10	0.985	1.000	0.985	CRTS	2	0	0	0	0	0	D50	35	0.948	0.952	0.995	IRTS	0	4	5	5	0	0
D12	11	0.984	1.000	0.984	CRTS	14	0	0	0	0	0	D7	36	0.947	0.948	0.999	DRTS	0	4	5.5	5.5	0	0
D24	12	0.983	0.987	0.996	IRTS	0	3	1.3	1.3	0	−0.6	D18	37	0.943	0.953	0.989	IRTS	0	3	6.8	4.9	0	0
D5	13	0.98	0.981	0.999	DRTS	0	4	1.9	1.9	0	0	D34	38	0.942	0.947	0.995	IRTS	0	3	5.7	5.6	0	0
D31	14	0.978	0.985	0.994	DRTS	0	4	1.6	1.6	0	0	D27	39	0.942	0.951	0.991	IRTS	0	4	5.2	5.2	0	0
D10	15	0.978	1.000	0.978	CRTS	10	0	0	0	0	0	D32	40	0.941	0.952	0.989	IRTS	0	3	5.1	5.1	0	−0.2
D2	16	0.973	0.976	0.997	DRTS	0	3	2.5	2.5	0	−2.8	D35	41	0.941	0.954	0.986	DRTS	0	3	4.8	9.1	0	0
D20	17	0.971	0.972	0.999	DRTS	0	4	2.9	2.9	0	0	D43	42	0.94	0.957	0.982	IRTS	0	2	6.6	4.5	0	−0.7
D42	18	0.971	0.977	0.994	DRTS	0	3	2.4	2.4	−0.2	0	D39	43	0.938	0.945	0.993	IRTS	0	4	5.8	5.8	0	0
D11	19	0.969	0.992	0.977	DRTS	0	2	3	0.8	0	−9.9	D36	44	0.938	0.94	0.998	IRTS	0	3	6.4	7.5	0	0
D46	20	0.968	0.973	0.995	IRTS	0	4	2.8	2.8	0	0	D26	45	0.934	0.968	0.964	IRTS	0	2	8.2	3.3	−1.5	0
D3	21	0.968	0.994	0.973	IRTS	0	2	0.6	2.3	−6.4	0	D33	46	0.932	0.933	0.999	IRTS	0	4	7.2	7.2	0	0
D21	22	0.967	0.99	0.977	IRTS	0	2	4.3	1	0	−4.5	D8	47	0.926	0.939	0.986	IRTS	0	2	6.8	6.5	0	−2.4
D4	23	0.967	0.969	0.998	DRTS	0	3	3.2	9.2	0	0	D13	48	0.913	0.92	0.993	DRTS	0	3	8.7	11.6	0	0
D40	24	0.966	0.967	0.999	DRTS	0	4	3.4	3.4	0	0	D14	49	0.912	0.913	0.999	IRTS	0	4	9.6	9.6	0	0
D1	25	0.965	0.98	0.985	IRTS	0	3	2	2.1	0	0	D23	50	0.906	0.928	0.977	IRTS	0	3	7.8	7.8	0	−2.2
		0.982	0.989	0.993				2.36	2.53					0.941	0.953	0.988				5.44	5.88		

Notes: 1. "Refs" denotes the number of times the efficient DMUs are referred to by the inefficient DMUs. 2. "Peers" denotes the number of times the inefficient DMUs are referring to other efficient DMUs. 3. The values of BCC, Scale, RTS, Refs, Peers, and room for improvement are obtained under BCC model.

3.2. Reference indicators of DMUs

The column "Refs" indicates the number of times the efficient DMUs referred by the inefficient DMUs. Table 2 indicates that a total of 11 DMUs with the pure technical efficiency (BCC score) equal to 1 have Refs values due to their excellent performance: no improvement in input and output items is needed for them. For example, D37 is the DMU most referred to and performs best: there are 25 inefficient DMUs referring to it. However, there is only 1 inefficient DMU referring to D49. Because D37's actual values (students' ratings of teachers) of output and input items (O1 = 4.05, O2 = 4.15, I1 = 3.9, and I2 = 3.95) are superior to those of D49 (O1 = 2.95, O2 = 3.6, I1 = 3.3, and I2 = 3.05). D37 belongs to the high actual value group with more DMUs; D49 belongs to the lower actual value group with fewer DMUs. It means that the key factors of the DMUs with higher actual values like the preparation of teaching contents, teaching skills, fair grading, and learning performance are more recognized by students. As for those DMUs with the BCC score inferior to 1 and not being located on the pure technical efficiency frontier curve, no DMUs will refer to them; this explains why their Refs values are all equal to 0.

The column "Peers" indicates the number of times the inefficient DMUs refer to other efficient DMUs on the efficient frontier curve in the same range of input and output items' actual values. The DMUs with a BCC score inferior to 1 are those with lower performance in the pure technical efficiency. For example, D11, ranked 19 refers 2 times to other efficient DMUs.

3.3. Room for improvement analysis

Using DEA can provide concrete and practical strategies for inefficient DMUs. The efficient frontier curve analyzes how many efforts and how much room for improvement is necessary for the output performance to come close to the efficient frontier and to reduce the gap between the actual output performance and the targeted output performance.

In Table 2, the column "Room for improvement" indicates how much improvement is needed for the DMUs and what items are concerned. The room for improvement of the output oriented BCC model emphasizes on how much the insufficiency of the output performance is under the current input resources without additional input efforts. It explains why I1 and I2, the values of the input items are mostly 0. That is, most DMUs' I1 and I2 need no further improvement; they focus only on the output side. Moreover, all the values in the columns O1 and O2 are positive: additional efforts on the output side are necessary in order to increase the output performance. The average values of O1 and O2 for the inefficient DMUs of top 50% are 2.36 and 2.53, respectively; those of the other inefficient DMUs are 5.44 and 5.88, respectively. This means that the inefficient DMUs of top 50% need less improvement in fair grading and students' learning performance than the other 50% inefficient DMUs. Furthermore, O2 needs firstly to be improved. That is, learning performance should be improved prior to fair grading.

D23 (ranked 50) has the lowest overall technical efficiency of 0.906. There is still a 7.8% effort to do in fair grading and students' learning performance. In this case, teachers are suggested to announce grading criteria clearly at the beginning of the semester to help students obtain guidance before preparing for exams and acquire the knowledge necessary to cover the writing course's important topics; it will help increase the value of O2 and enhance students' learning performance at the same time.

The values in the columns I1 and I2 are sometimes negative for some of the DMUs; it means that these DMUs are recommended to reduce the inputs so as to enhance the overall efficiency. Table 2 reveals that in the I1 column, only the values of D42, D3, D28, D48 and D26 are different from zero. It indicates that the preparation of teaching materials and course contents are too much and too complicated for students to assimilate. Teachers should lightly reduce the quantity of course contents of indicated percentage in order to improve students' learning performance. Teachers need to make extra efforts to teach more than just grammar and syntax. Writing courses are not only to correct language mistakes, but also to acquire a fair understanding of Western culture combined with a sharp logic. A teacher with writing experience and a good understanding of Western culture will be more suitable for this type of course. As for the case of inefficient DMUs with negative values in I2, it means that teachers over explain the writing course contents. In order to enhance students' learning performance, teachers are advised to give students more opportunities to do exercises or to revise the course contents.

3.4. Importance of efficiency values in managerial strategy

DEA includes the function and concept of benchmarking, that is, the process of measuring performance using specific indicators (I1, I2, and O1, O2) resulting in quantitative values compared with others. The inefficient DMUs can refer to the outstanding DMUs in their range and are not always compared with the ones with the highest range of actual values. The calculated results change our original viewpoint and reveal that some DMUs with higher actual values (students' ratings of teachers) of input and output items have lower efficiency as demonstrated in Table 3. This phenomenon is valid for each item. Hence, Table 3 only lists the values of output item O2 as an explanatory case and includes efficiency values, actual values, and the correspondent different facet reference sets of D11, D28, and D23. D11 is a representative of the DMUs with higher ranking but lower actual values of students' ratings. D28 and D23, on behalf of the DMUs with lower ranking but higher actual values of students' ratings, refer to the efficient DMUs in the same higher range. In fact, the CCR model or the BCC model estimate the efficiency frontier by the ratio of linear combination of inputs divided by the linear combination of outputs and measure the relative efficiency of each DMU. This is the reason why even the DMUs with lower actual values of input and output items may obtain relative efficiency values equal to 1. These DMUs become facet reference sets, which form efficiency frontier curves. The

Table 3
Some indicative values of 3 inefficient DMUs in different facet reference sets.

DMU	CCR (ranking)	BCC	Actual value	Target value	Room for improvement (%)	Reference set	Reference set's actual value	Reference set's contributions (%)
D11	0.969 (19)	0.992	3.30	3.33	0.80	D10	2.85	10.08
						D29	3.39	89.92
D28	0.951 (31)	0.994	4.55	4.7	3.25	D12	4.68	63.64
						D45	4.73	36.36
D23	0.906 (50)	0.928	4.36	4.7	7.78	D12	4.68	50.85
						D37	4.15	0.78
						D45	4.73	48.38

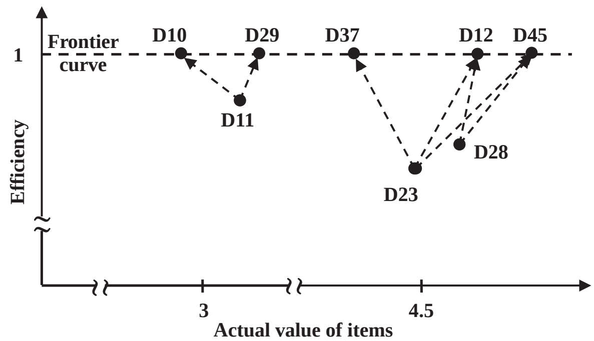

Fig. 1. Referring diagram of inefficiency DMUs to efficient DMUs.

frontier curves in 2-dimension (dotted line) are illustrated in Fig. 1, which is a diagram of the inefficient DMUs referring to the efficient DMUs. The symbols "≈" on Fig. 1 mean that the axes are not to scale. The DMUs with lower actual values of input and output items refer to the DMUs on the efficiency frontier curve in the same range and near them. The efficiency is calculated in accordance with their distance to the curve. As a result, if a DMU is located very near the efficiency frontier curve, it will obtain an efficiency value very close to 1, even with a lower actual value. For example, as illustrated in Fig. 1, D11 is among the best of a group with lower actual values and refers to two efficient DMUs of the same range, D10 and D29. The reference set's contribution of D29 (89.92%) to D11 is much higher than that of D10 (10.08%) because D11's actual value of O2 is close to that of D29. D28 and D23 belong to a group with higher actual values and have two efficient DMUs in common, D12 and D45. Since D28's location is closer to the efficiency frontier curve than D23, its efficiency is higher than D23's.

The fact that the DMUs with higher actual values of input and output items obtain higher efficiency values is only valid when they refer to the same reference set like in the case of D23 and D28. As shown in Table 3, D11's actual value and those of its reference set (D10, D29) are around 3, which belong to the range of lower values; D28 and D23's actual values and those of their reference sets [(D12, D45) and (D12, D37, D45)] are around 4.5, which belong to the range of higher values. Even though D11 has a lower actual value than those of D28 and D23, it is ranked higher among the 50 DMUs in our paper. This phenomenon confirms perfectly our analysis mentioned above and is very encouraging in the managerial implications in the field of education. Because the efficiency calculated by using the DEA model helps to know how much improvement needed is for the inefficient DMUs. This encourages the inefficient DMUs to always compare themselves with the efficient DMUs in their range. If these DMUs are compared with the efficient DMUs of the highest range, their motivation will be probably shattered and they will give up. Hence, the inefficient DMUs with lower actual values may endeavor to make improvement little by little. Students, who have different characteristics, cannot always acquire the same knowledge and learning results from a same teacher even if they make the same amount of efforts. Teachers are suggested to ponder the proportion of different grading items in students' final grades such as their attendance, learning attitude, and participation during the class. As for the inefficient DMUs with high actual values, the results of the DEA model may encourage them to aspire to a higher target and to guide them to make continuous progress. These results may give DMUs clear indicators and the percentage of which input and output items to improve.

Moreover, we find that the benchmarking characteristics of the DEA model can be used to automatically segment all the DMUs into different levels based on their actual values of input and output items. The efficient DMUs on the frontier curve can be considered as the boundaries of these levels. These boundaries are systematically defined by the DEA model according to the statistic distribution of all the DMUs and may change according to the indicators fed into the performance evaluation mechanism. Thus, the subjective judgments and errors made by human beings can be avoided to some extent. For example, the DMUs [D10, D29, and D11] form a group and belong to level 1 due to their lowest scores in terms of actual values compared with those of all the DMUs; D10 and D29 are the floor and the ceiling of level 1 respectively. The DMUs [D37, D12, D45, D23 and D28] form another group and belong to the highest level due to their highest scores in terms of actual values; D37 and D45 are the floor and the ceiling of the highest level respectively. The advantage of this classification is not only that the inefficient DMUs can obtain clear indication of improvement, make progress step by step, and reach the frontier curve during the benchmarking process, but also that the efficient DMUs on the frontier curve and belonging to lower levels can refer to higher levels and aspire to a higher target instead of being stuck in the bottlenecks. With time, we can expect all the DMUs to become outstanding and to belong to the highest level. The directors of educational institutions or corporations can refer to this concrete classification to give fairer rewards and encourage DMUs to always perform better.

4. Conclusions and implications

The paper uses DEA, a reliable and robust evaluation method to explore the key indicators contributing to students' learning performance for English freshmen writing courses in a university of Taiwan. The paper aims at providing DMUs with objective and impartial measuring indices under limited teaching and learning resources.

We observe that 10% of the DMUs have an overall technical efficiency value (CCR score) of 1 and do not need any improvement in the inputs or in the outputs because they have reached their optimal state, meaning that both teachers and students feel at ease and are motivated to work. This state requires a good atmosphere in the class and good teaching preparation and skills from the part of teachers. Students have to show their efforts during the training and accept criticism. However, if they are over criticized by their teachers, they will probably loose their motivation.

Thirty-nine inefficient DMUs have the pure technical efficiency values (BCC score) smaller than 1. Among which, six have pure technical efficiency values greater than scale efficiency values: students are unable to assimilate the entire contents of the course. The other 33 inefficient DMUs have pure technical efficiency values smaller than scale efficiency values: some students think that their teacher has not enough professional knowledge and experience to teach a course of English writing.

The results of the DEA model applied in learning performance change our original viewpoint and reveal that some DMUs with higher actual values of inputs and outputs have lower efficiency because the relative efficiency of each DMU is measured by their distance to the efficiency frontier. DMUs may refer to different facet reference sets according to their actual values located in lower or higher ranges. In the managerial strategy of educational field, the paper can encourage the inefficient DMUs to always compare themselves with the efficient DMUs in their range and to make improvement little by little. The results of the DEA model can also give clear indicators and the percentage of which input and output items to improve.

The paper demonstrates that the benchmarking characteristics of the DEA model can automatically segment all the DMUs into different levels based on their actual values of input and output items. Moreover, the efficient DMUs on the frontier curve can be considered as the boundaries of the classification which are systematically defined by the DEA model according to the statistic distribution of all the DMUs and may change according to the indicators fed into the performance evaluation mechanism. This concrete classification and the evaluation approach of learning performance can be employed to the field of foreign language learning as well as other branches of learning or even to corporate employee training.

References

Ahn, T., Arnold, V., Charnes, A., & Cooper, W. W. (1989). DEA and ratio efficiency analyses for public institutions of higher learning in Texas. *Research in Governmental and Nonprofit Accounting, 5*, 165–185.

Arreola, R. A. (1986). Evaluating the dimensions of teaching. *Instructional Evaluation, 8*(2), 4–12.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science, 30*(9), 1078–1092.

Barcelos, A. M. F., & Kalaja, P. (2003). Conclusion: Exploring possibilities for future research on beliefs about SLA. In P. Kalaja & A. M. F. Barcelos (Eds.), *Beliefs about SLA: New research approaches* (pp. 231–238). Amsterdam: Kluwer Acadamic.

Boland, D. L., & Sims, S. L. (1988). A comprehensive approach to faculty evaluation. *Journal of Nursing Education, 27*(8), 354–358.

Chacko, T. I. (1983). Student ratings of instruction: A function of grading standards. *Educational Research Quarterly, 8*(2), 19–25.

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

Clinton, B. D., & Kohlmeyer, J. M. (2005). The effects of group quizzes on performance and motivation to learn: Two experiments in cooperative learning. *Journal of Accounting Education, 23*(2), 96–116.

Colbert, A., Levary, R., & Shaner, M. (2000). Determining the relative efficiency of MBA programs using DEA. *European Journal of Operational Research, 125*, 656–660.

Dickinson, D. J. (1990). The relationship between ratings of teacher performance and student learning. *Contemporary Educational Psychology, 15*(2), 142–151.

Fu, T. T., & Huang, M. Y. (2009). Performance ranking and management efficiency in colleges of business: A study at the department level in taiwanese universities. In J. D. Lee & A. Heshmati (Eds.), *Productivity, efficiency, and economic growth in the Asia-Pacific region*. *Contributions to Economics* (pp. 197–215). Heidelberg: Physica-Verlag.

Glass, J. C., Mckillop, D. G., & O'Roruke, G. (1998). A cost indirect evaluation of productivity change in UK university. *Journal of Productivity Analysis, 10*(2), 153–175.

Haskell, R. E. (1997). Academic freedom, tenure and student evaluation of faculty: Galloping polls in the 21st century. *Education Policy Analysis Archives*, 5 (6). Retrieved March 22, 2010, from <http://epaa.asu.edu/epaa/v5n6.html>.

Hsu, M. H. (2010). Proposing an interactive speaking improvement system for EFL learners. *Expert Systems with Applications, 37*(1), 414–418.

Johnes, G., & Johnes, J. (1993). Measuring the research performance of UK economics departments: An application of data envelopment analysis. *Oxford Economic Papers, 45*(2), 332–347.

Lee, C. C. (2009). Analysis of overall technical efficiency, pure technical efficiency and scale efficiency in the medium-sized audit firms. *Expert Systems with Applications, 36*(8), 11156–11171.

Lin, H. T. (2009). Efficiency measurement and ranking of the tutorial system using IDEA. *Expert Systems with Applications, 36*(8), 11233–11239.

Lin, T. T., Lee, C. C., & Chiu, T. F. (2009). Application of DEA in analyzing a bank's operating performance. *Expert Systems with Applications, 36*(5), 8883–8891.

Madden, G., Savage, S., & Kemp, S. (1997). Measuring Public Sector Efficiency: A Study of Economics Department at Australian Universities. *Education Economics, 5*(2), 153–168.

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist, 52*(11), 1187–1197.

McGowan, W. R., & Graham, C. R. (2009). Factors contributing to improved teaching performance. *Innovative Higher Education, 34*(3), 161–171.

Ministry of Education (2001). *The Republic of China Yearbook 2001*. Chapter 17. Education. Retrieved March 22, 2010, from <http://www.gio.gov.tw/taiwan-website/5-gp/yearbook/2001/chpt17.htm>.

Ministry of Education (2009a). *The Republic of China Yearbook 2009*. Chapter 17. Education. Retrieved March 22, 2010, from <http://www.gio.gov.tw/taiwan-website/5-gp/yearbook/ch17.html>.

Ministry of Education (2009b). *The Republic of China Yearbook 2009*. Chapter 2. Language and People. Retrieved March 22, 2010, from <http://www.gio.gov.tw/taiwan-website/5-gp/yearbook/ch02.html#DemographicTrends>.

Ministry of Interior (2010). *Sixth Weekly Report, Statistical bulletin 2010*. Retrieved March 22, 2010, from <http://www.moi.gov.tw/stat/news_content.aspx?sn=3829>.

Mitzel, H. E. (1982). Teacher effectiveness. In *Encyclopedias of educational research*. New York: McMillan Publishing Co.

Ryan, J. J., Anderson, J. A., & Birchler, A. B. (1980). Student evaluations: The faculty responds. *Research in Higher Education, 12*(4), 317–333.

Samoilenko, S., & Osei-Bryson, K. M. (2008). Increasing the discriminatory power of DEA in the presence of the sample heterogeneity with cluster analysis and decision trees. *Expert Systems with Applications, 34*(2), 1568–1581.