Teaching and Teacher Education 28 (2012) 382–395

Contents lists available at SciVerse ScienceDirect

Teaching and Teacher Education

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

Application of data envelopment analysis on the indicators contributing to learning and teaching performance

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

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

A R T I C L E I N F O

Article history:
Received 8 February 2011
Received in revised form
27 June 2011
Accepted 18 November 2011

Keywords:
Data enveloped analysis (DEA)
Key performance indicators
Teaching performance improvement

A B S T R A C T

This paper applies data envelopment analysis (DEA) to explore the quantitative relative efficiency of 18 classes of freshmen students studying a course of English conversation in a university of Taiwan from the academic year 2004–2006. A diagram of teaching performance improvement mechanism is designed to identify key performance indicators for evaluation in order to help teachers concentrate their efforts on the formulated teaching suggestions. The sensitivity study highlights the priority of the richness of course contents over the other evaluated indicators. The performance improvement mechanism can help decision-makers to design educational policies.

1. Introduction

Over the last two decades, the number of colleges and universities in Taiwan has increased by 40 from 123 in 1991 to 163 in 2010 (Ministry of Education in Taiwan, 2011a). According to the Ministry of Interior in Taiwan (2010), the country has the lowest birth rate in the world: 0.83. As a result, the number of students will probably decrease dramatically in the coming years even though the acceptance rate for colleges and universities has reached 94.87% in 2010 (Ministry of Education in Taiwan, 2011b). In Taiwan, higher education has become increasingly competitive and universities must enhance their reputation to ensure their future. In recent years, in order to allocate more efficiently limited education resources and control the quality of schools, the Ministry of Education is undertaking a performance evaluation for all colleges and universities. This is also done in many other countries around the world. Higher education institutions (HEIs), especially the private ones, hope to obtain an excellent evaluation in order to receive more financial support from the Ministry of Education and to avoid low student enrollment, high graduate unemployment, credential inflation, and even university closure. Taiwan is not the only country to face this kind of situation. In Europe, many countries have a very low birth rate as well. For example, García-Aracil and Palomares-Montero (2008) note that the current decline in student numbers is extremely important in Spain. The question of the survival of many educational institutions is inevitably posed. The gradual decrease in some European countries has motivated governments to implement strategies to measure universities' performance. In Japan, according to Burden (2008), as the numbers of prospective students decline and as most universities depend on student tuition fees for survival, students are becoming courted customers and student evaluation of teaching surveys (SETs) have become indispensable to uphold the quality of university education.

The Ministry of Education in Taiwan promotes an Excellent Teaching Project. This program aims at encouraging HEIs to improve teaching effectiveness. The universities which obtain this project generally set up an Excellent Teaching Center and define some indicators of teaching performance for teachers to follow. The indicators selected for this paper satisfy teaching effectiveness not only at the level of English conversation courses or any other specific subject, but also at the level of the Excellent Teaching Centers of Higher Education institutions as well as at the level of the Ministry of Education. The design of teaching performance indicators follows this bottom-up program in order to enforce the competitiveness of the country. The results of the performance evaluation via the selected indicators can serve as a reference for the Ministry of Education to formulate educational policies.

Among the many evaluated items by the Ministry of Education, one of the most fundamental is students' learning performance. The quality of students' learning has an influence not only on the

* Corresponding author. Tel.: +886 4 26328001x12021; fax: +886 4 26334712.
E-mail addresses: bmon@pu.edu.tw (B. Montoneri), tjlin@mail.ndhu.edu.tw (T.T. Lin), cclee.vera@msa.hinet.net (C.-C. Lee), slhuangmon@gmail.com (S.-L. Huang).

0742-051X/$ – see front matter
doi:10.1016/j.tate.2011.11.006

long-term relationship of a country's future growth and competitiveness, but also on the short-term impact, on the employment rate of graduate students, or even on whether the school can attract enough freshmen. The evaluation of learning performance is also used for the organizations possessing human resources as their major assets. In a fierce competitive market, employees' working performance and learning performance can be as fatal as in other industries and fields.

Key performance indicators (KPIs) are measures of accomplishment, tools used by individuals and organizations to track their progress and success. Higher education is now seen as an economic commodity, leading to a greater interest from governments and funding agencies in measuring the employability of students through measures of learning and their employment outcomes (Australian Learning and Teaching Council, 2008). According to Azma (2010), without the evaluation of performance based on key factors and indicators, there will be no permanent change and improvement in the enhancement of the quality of the universities. In order to find out which indicators have the biggest impact on students' learning performance, this paper adopts the data envelopment analysis (DEA) methodology, an evaluating model able to provide a comparative efficiency indicator of the units to be evaluated (Martin, 2006). As mentioned in the literature review below, the DEA evaluation method has been widely applied in various industries in Taiwan and abroad since 1978 and is proved to be quite reliable. Even though it has also been used to assess the efficiency of HEIs, there is relatively little application of DEA on language learning performance which is mostly analyzed by qualitative methods. The research object consists of 18 classes of freshmen students taught by full-time teachers from a department of English in a university of Taiwan, entering from the academic year 2004, 2005, and 2006. These classes, which follow the same training program of English conversation for one semester, are selected as the evaluated units.

This paper proposes a method to find out which indicators of teaching and learning performance are the most suitable. To do so, we focus on 4 indicators as an example: two inputs (the richness of course contents, the diversity of accessed multiple teaching channels) and two outputs (the positive degree of teaching attitude and students' learning performance). These 4 indicators were selected among a total of 10 by using SPSS. The research results acquired by applying DEA are expected to indicate whether the existing teaching scale is in optimal size and whether students' learning efforts and teachers' teaching efforts will reach the expected performance. The sensitivity study of the evaluated indicators, which consists in withdrawing respectively one of the 4 indicators studied in this paper, is performed to show the priority of the evaluated indicators and observe their impacts on each evaluated unit's efficiency score.

The remainder of this paper is organized as follows: Section 2 (literature review) presents a few academic researches in relation with our paper. Section 3 (methodology and chosen evaluated indicators) designs a diagram of teaching performance improvement mechanism and explains the DEA method and the important indicators discussed in this paper. Section 4 (empirical results and suggestions) presents the obtained numerical results based on the empirical data which include the efficiency analysis and sensitivity study. Section 5 describes the main findings, limitations and directions of future studies.

2. Literature review

Based on the purposes above, this study probes into literature related to the application of data envelopment analysis in education around the world and to the indicators contributing to teaching and learning performance in HEIs.

2.1. Origins and application of DEA

Data envelopment analysis (DEA) is a quantitative method which can assess the relative efficiency of evaluated units, commonly called decision making units (DMUs) within a sample (Samoilenko & Osei-Bryson, 2008). According to the suggestions provided by DEA, the inefficient DMUs can refer to the outstanding DMUs to effectively reach an efficient state. Førsund and Sarafoglou (2002) review the literature concerning DEA and notably discuss its origins and further development. The starting point is generally attributed to Farrell's seminal 1957 paper on concepts of efficiency. Originally, in Farrell (1957), the concept of efficiency measurement was restricted to a single output and multiple inputs in the development of econometric techniques. Since 1978, DEA includes the function and concept of benchmarking and has been applied in various fields. The concept was introduced in the journal literature by the highly influential paper of Charnes, Cooper, and Rhodes (1978); they notably expanded Farrell's efficiency measurement concept to the concept of multiple inputs and multiple outputs. Their method, called the "Charnes-Cooper-Rhodes (CCR) model" or "CCR model". The CCR (ratio) model is the most famous and the most widely used DEA model nowadays. It estimates the efficiency frontier by the ratio of inputs' linear combinations to outputs' linear combinations in order to measure the relative efficiency of each DMU. Charnes et al. (1994) elucidated that CCR used the optimization method of mathematical programming to perform the efficiency calculation and to demonstrate the multiple-outputs/multiple inputs case. CCR started a new active research field, popularly called DEA. According to Lin, Lee, and Chiu (2009) and Lee (2009), the efficiency value of the CCR model corresponds to the overall efficiency of an evaluated unit. If the efficiency value equals 1, the evaluated unit is efficient (of optimal performance); if the efficiency value is less than 1, the evaluated unit needs some improvement.

2.1.1. Application of DEA in various fields

DEA is a reliable and robust evaluation method. It has been applied in various industries such as electricity sector (Cherchye & Post, 2003), high-tech industry (Kozmetsky & Yue, 1998; Lai, 2007; Thore et al., 1996), financial industry (Jemric and Vujcic, 2002; Lin et al., 2009), medical industry (Valdmanis, 1990), and transport industry (Yang, 2005).

2.1.2. Application of DEA in education around the world

Data envelopment analysis (DEA) has also been used to assess the efficiency of HEIs: Ahn et al. (1989) on US universities during 1981–1985; Glass, Mckillop, and O'Roruke (1998) and Johnes and Johnes (1993) on UK universities; Ng and Li (2000) in China; McMillan and Datta (1998) in Canada; and Avkiran (2001) in Australia. An assortment of methodological approaches has been employed in an effort to resolve the problem of efficiency measurement from the early studies which use ordinary least-squares (OLS) regression methods (Johnes & Taylor, 1990) to the more recent studies which use frontier methods such as DEA (Johnes, 2006). Abbott and Doucouliagos (2003) determine the teaching and research performance of Australian public universities in 1995. However, fewer studies measure the efficiency at the departmental level (Madden, Savage, & Kemp, 1997 on economics departments in Australian universities; Johnes & Johnes, 1993 on economics departments in the UK in 1984–1988; Colbert, Levary, & Shaner, 2000, on MBA programs in the US). DEA was also applied in Spain to analyze the research and teaching activity of departments in public universities (García Valderrama, 1996; Pina and Torres, 1995).

2.2. The key performance indicators of evaluation

Key performance indicators (KPIs) are tools used by individuals and organizations to track their progress and success. They are also used by governments to assess the efficiency of HEIs. However, the ambiguity found in education performance measurements makes it relatively difficult to capture the interaction among the various inputs and outputs and the limitations with the selected output specification (García-Aracil & Palomares-Montero, 2008; Joumady & Ris, 2005). For example, the undergraduate student number or doctoral student number can be both a teaching and a research indicator (García-Aracil, 2006). Various studies have been conducted on the key performance indicators of evaluation, but there is little consensus concerning the choice of indicators to assess the performance of HEIs. Therefore, our demonstration on how to screen primary indicators will be quite useful for further studies in different countries or fields. The following literature reviews some studies conducted in this field according to their geographical location. A recapitulation of some key performance indicators studies of evaluation in global HEIs are listed in Table 1 – Table 3.

2.2.1. The key performance indicators of evaluation in America and Australia

The higher education sector in the USA is a state based system; the U.S. Department of Education acts mainly as a repository of federal funds, and the quality assurance of post-secondary education is delegated to the states and accreditation bodies. According to Australian Learning and Teaching Council (2008), there are calls for greater government oversight of HEIs in the US through the use of standardized indicators and measures.

Wolf, Bender, Beitz, Wieland and Vito (2004) describe in La Salle University School of Nursing, Philadelphia (USA), the strengths (being a knowledgeable and strategic teacher, creating a positive learning environment, demonstrating professionalism, demonstrating positive personal traits, and displaying scholarly traits) and weaknesses (providing poor delivery of course contents, acting disorganized, being inaccessible, and displaying weak teaching skills) in faculty teaching performance as reported by undergraduate and graduate nursing students.

McGowan and Graham (2009) study the factors contributing to improved teaching at Brigham Young University (BYU), a private church-sponsored university. According to their research, the four indicators 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.

According the Australian Learning and Teaching Council (2008), Australia has been one of the global leaders in development of processes for quality assurance in teaching and learning and in seeking valid and reliable indicators for performance. In this huge report of 125 pages, many indicators are presented such as perceived teaching quality, course experience questionnaire, student engagement, progress rate, retention rate, graduate full-time employment, graduate part or full-time further study, graduate satisfaction with generic skills obtained, graduate satisfaction with the quality of teaching received, and graduate overall satisfaction. Chalmers (2008) presents the four main types of performance indicators, that is, input, process, output, and outcome. Chalmers & Thomson (2008), in the national survey of Australian practice, highlight the widespread use of process indicators (measuring the quality of delivery of educational programs, activities and services) in Australian universities and identify thirteen different categories.

2.2.2. The key performance indicators of evaluation in Europe

29 European countries signed the Bologna Declaration in 1999 to increase the cooperation and the development of compatible and

Table 1
A short survey of various KPIs in some research papers in America and Australia.

Author	Performance indicators	Main findings	Location
Wolf et al. (2004)	Performance strengths: 1. Being a knowledgeable and strategic teacher 2. creating a positive learning environment 3. demonstrating professionalism 4. displaying scholarly traits 5. being supportive Performance weaknesses: 1. Poor delivery of content 2. acting disorganized 3. being inaccessible 4. displaying weak teaching skills; being dishonorable 5. being unprofessional 6. displaying negative traits	1. undergraduate students considered teaching strategies, faculty enthusiasm and knowledge base, and faculty support of student efforts as strengths of faculty teaching performance 2. faculty patience was valued, as were faculty answers to their questions 3. graduate students' comments emphasized faculty excellence, knowledge base, and flexibility, and also emphasized a caring approach	USA
McGowan and Graham (2009)	1. active/practical learning (real-world experiences and in-class discussions) 2. teacher/student interactions (knowing students personally) 3. clear expectations/learning outcomes 4. faculty preparation	1. 19% of the BYU faculty members showed significant improvement in their teaching over a 3-year period 2. improvement attributed to following a few simple steps such as active and practical learning experiences, seeking to have meaningful interactions with students	USA
Australian Learning and Teaching Council (2008)	1. perceived teaching quality 2. course experience questionnaire 3. student engagement 4. progress rate 5. retention rate 6. graduate full-time employment 7. graduate part or full-time further study 8. graduate satisfaction with generic skills obtained 9. graduate satisfaction with the quality of teaching received 10. graduate overall satisfaction	1. Australia uses a performance budgeting model rather than performance funding as do New Zealand, the United Kingdom and Japan 2. each indicator is not strictly linked to funding and is only a factor in determining the total allocation of funds 3. the trend towards quality improvement is global and has not been limited to higher education	Australia

Table 2
A short survey of various KPIs in some research papers in Spain and UK.

Author	Performance indicators	Main findings	Location
Martin (2006)	1. human resources 2. financial resources 3. material resources 4. credits registered × experimental coefficient 5. Ph.D. credits offered 6. Ph.D. completions 7. annual research incomes 8. scientific production index	1. shortcomings of the Spanish public university system, rigidity of the functional structures 2. 10 departments on 52 studied obtain an evaluation of 100% 3. choice of discipline by students conditioned by the job prospects it generates	Spain
García-Aracil and Palomares-Montero (2008)	1. total expenditure 2. academic staff 3. non-academic staff 4. number of graduates 5. publications 6. total amount of applied research	1. analyze the productivity change of the Spanish public universities from 2002 to 2004 2. most productivity growth is associated with improvements in research and knowledge transfer than teaching 3. annual productivity growth in research model largely attributable to efficiency improvements rather than technological progress	Spain
Johnes (2006)	Some of the indicators: 1. score based on best 3 A levels or equivalent 2. gender 3. school 4. % of graduates who are female 5. % of graduates who did not attend an independent school 6. pass/other	1. efficiencies derived from DEAs performed at an aggregate level include both institution and individual components, and are misleading 2. measures of the efficiency of departments derived from individuals' efficiencies are much more highly correlated with department level efficiency scores	UK

Table 3
A short survey of various KPIs in some research papers in China and Taiwan.

Author	Performance indicators	Main findings	Location
Zhou and Wang (2009)	1. teacher as labor power index 2. financial power (including state appropriation, tuition income, self-financing) 3. physical power (teaching building, lab equipment, books) 4. number of graduates 5. scientific research and publications	1. build a scientific evaluation index system 2. discuss the quality of input & output efficiency with DEA 3. highlight that 56% of the studied universities are not efficient	China
Wu and Li (2009)	1. financial 2. customer 3. internal process 4. learning & growth	1. calculate the relative efficiency between 15 Science and Technology universities	China
Wu (2010)	1. clients 2. operations 3. resources, participants 4. services	1. calculate the relative efficiency between 24 Science and Technology universities in China 2. propose a model aimed at improving Chinese university performance ranking	China
Fu and Huang (2009)	1. average monthly starting salary of graduates 2. average search duration of graduates for the first job 3. average monthly current salary of graduates 4. student satisfaction with quality of curriculum in major field 5. student satisfaction with quality of curriculum in non-major fields	1. evaluate the performance of departments from the point of view of prospective students and recruiters 2. concerning job market performance the average score for the overall sample is 81.29% 3. the departments in public schools perform better than those in private schools in the job market	Taiwan
Montoneri, Lee, Lin, and Huang (2010a)	1. preparation of teaching contents 2. teaching skills 3. fair grading 4. students' learning performance	1. teachers should announce the grading criteria early 2. teachers' over explanation of course contents can decrease students' concentration during the class	Taiwan
Montoneri, Lee, Lin, and Huang (2010b)	1. richness of course content 2. diversity of accessed multiple learning channels 3. teaching attitude 4. students' learning performance	1. empirical results can identify the efficient classes and those needing some adjustments 2. indicators' improvement degree for each evaluated unit is analyzed 3. teachers are suggested to respond to students' questions more positively, to lightly reduce the quantity of teaching materials, and to improve the learning channels 4. experienced teachers are suggested to help their colleagues and to give them advices and ideas on how to make their class more attractive	Taiwan
Montoneri, Lee, Lin, and Huang (2011)	1. preparation of teaching contents 2. teaching skills 3. fair grading 4. students' learning performance	1. some evaluated classes with higher actual values of inputs and outputs have lower efficiency because the relative efficiency of each evaluated class is measured by their distance to the efficiency frontier 2. the benchmarking characteristics of the DEA model can automatically segment all the evaluated classes into different levels based on the indicators fed into the performance evaluation mechanism	Taiwan

comparable higher education systems. The main objectives are to promote transparency, mobility, employability, and student-centered learning within Europe (Bologna Declaration, 1999).

Martin (2006) applies DEA methodology to assess the performance of the departments of the University of Zaragoza (Spain); the selected indicators concern both the teaching and the research activity of the studied departments in the academic year 1999. The inputs are human resources, financial resources and material resources; the outputs are credits registered × experimental coefficient, Ph.D. credits offered, Ph.D. completions, annual research incomes, and scientific production index. Martin (2006) defines four models with different combinations of inputs and outputs. As a result, 10 departments on 52 studied obtain an evaluation of 100% with all four models. The sciences knowledge area within the University stands out for the research activity of its departments, but their teaching activities are weaker and the number of students enrolled is gradually diminishing.

García-Aracil and Palomares-Montero (2008) collect the data from 47 public institutions and apply the Malmquist non-parametric approach to analyze the productivity change of the Spanish public universities from 2002 to 2004. They use total expenditure, academic staff, and non-academic staff as inputs and number of graduates, publications, and the total amount of applied research as outputs. They find that most productivity growth is associated with improvements in research and knowledge transfer than teaching.

In the United Kingdom, the higher education system is highly managed and centralized. Although HEIs are self-governing and independent, most of them receive government funding. Like in many other countries, British policy makers increasingly desire to hold educational institutions accountable for their performance. According to Johnes (2006), there has been a growing need for accountability in sectors which are in receipt of large amounts of public money. He states that DEA is an attractive tool of analysis in the context of measuring the performance of HEIs and he applies this method to assess the teaching efficiency of 2547 Economics graduates from UK Universities in 1993. The multiple input multiple output nature of production in a HEIs makes DEA rather than stochastic frontier analysis (SFA) the ideal choice of technique in this context (Johnes, 2006). Using an output-oriented approach and indicators such as score based on best 3 A levels or equivalent, gender, school, % of graduates who are female, % of graduates who did not attend an independent school, and pass/other, he finds out that measures of the efficiency of departments derived from individuals' efficiencies are much more highly correlated with department level efficiency scores. Many other studies have measured efficiency measures in UK higher education, such as Tomkins and Green (1988), Johnes and Taylor (1990), Pugh, Coates and Adnett (2005) or Glass, McCallion, McKillop, Rasaratnam, and Stringer (2006).

2.2.3. The key performance indicators of evaluation in China and Taiwan

Zhou and Wang (2009) use DEA to analyze the efficiency of 16 universities in China. Their performance indicators are teachers as labor power index, financial power, physical power, number of graduates, and scientific research. They find out that 9 DMUs have reached the optimal state with a value of 1. The other DMUs are not considered efficient.

Wu and Li (2009) construct a performance measure indicators system for higher education using four perspectives: financial, customer, internal process, and learning & growth. They notably apply DEA to calculate the relative efficiency between 15 Science and Technology universities in China.

Wu (2010) uses PCA (Principal Component Analysis) and DEA methods to analyze five performance measure indicators for higher education, such as clients, operations, resources, participants, and services. This study concerns 24 Science and Technology universities in China and proposes a model aimed at improving Chinese university performance ranking.

Fu and Huang (2009) conducted a survey of college graduates in 2003 in Taiwan and collected different dimensions of performance indicators, including college graduate performance in the job market after graduation and student satisfaction with regard to the school environment and curriculum, as the student's devotion to the school and its related activities. Fu and Huang (2009) use an output-oriented BCC type of the DEA model to provide useful information for prospective students in terms of their choices regarding which college to join and to evaluate the relative resource use efficiency of schools for school administrators.

Montoneri, Lee, Lin, and Huang (2010a) explore the learning performance of English as Second Language (ESL) learners in a university of Taiwan. Freshmen following English writing courses from the academic year 2004–2006 are the research object. A total of 50 classes are selected as the evaluated units, named from D1 to D50. This paper uses students' ratings of teachers (questionnaires filled at the end of each semester) and applies data envelopment analysis (DEA) to calculate the evaluated classes' efficiency. In our study, four selected indicators satisfy the principle of Isotonicity and are proved to be representative after a correlation analysis. The two input indicators are the preparation of teaching contents and the teaching skills; they have a significant impact on the two output indicators, that is, the fair grading and the students' learning performance. Montoneri et al. (2010a) perform a preliminary qualitative analysis of the 4 indicators revealing that teachers should announce the grading criteria early in the semester to help students follow the rules and obtain guidance before preparing the exams. In addition, teachers' over explanation of writing course contents can also decrease students' concentration during the class: they are advised to give students more opportunities to do exercises or to revise the course contents so as to achieve the optimal performance of the outputs and enhance students' learning performance. It is suggested to conduct further quantitative studies of indicators' contributing to efficiency value in order to enhance the evaluated classes' performance.

Montoneri, Lee, Lin, and Huang (2010b) use the results of students' ratings of teachers for preliminary identification of the teaching resource inputs which have a significant impact on students' learning performance. Freshmen following English conversation courses from the academic year 2004–2006 are the research object. A total of 18 students' classes taught by full-time teachers are selected as the evaluated units. They are named from D1 to D18. The selected teaching resource inputs (the richness of course content and the diversity of accessed multiple teaching channels) show a very high degree of correlation with the selected outputs (the positive degree of teaching attitude and the students' learning performance). The empirical results can help to identify the evaluated classes with the best performance and those needing some adjustments. In addition, the indicators' improvement degree for each evaluated unit is also analyzed. Teachers are notably suggested to respond to students' questions more positively, to lightly reduce the quantity of teaching materials, and to improve the teaching channels (such as language learning websites, learning softwares, online courses). It is suggested that experienced teachers help their colleagues and give them suggestions and ideas on how to make their class more attractive.

Montoneri, Lee, Lin, and Huang (2011) apply DEA to assess the performance of English writing courses in a university of Taiwan and select the following indicators: preparation of teaching contents, teaching skills, fair grading, and students' learning performance. They propose an output-oriented model and demonstrate that some evaluated classes with higher actual values of inputs and outputs have lower efficiency because the relative efficiency of each evaluated class is measured by their distance to the efficiency frontier. Evaluated classes 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, this paper can encourage inefficient evaluated classes to always compare themselves with efficient evaluated classes in their range and make improvement little by little. This paper also demonstrates that the benchmarking characteristics of the DEA model can automatically segment all the evaluated classes into different levels based on the indicators fed into the performance evaluation mechanism. The efficient evaluated classes 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.

3. Methodology and chosen evaluated indicators

The efficiency assessment is often conducted by using the stochastic frontier analysis (SFA) or the data envelopment analysis (DEA). Which analysis is more suitable depends on the research environment (Ferrier & Knox Lovell, 1990). Once the defined function is inappropriate, the efficiency obtained by SFA will be inaccurate and meaningless. DEA uses linear programming analysis to calculate the relative efficiencies of evaluated units in order to avoid the evaluation bias due to the inappropriate function design.

There is a trend since the works of famous scholars such as Johnes and Johnes (1993) to use DEA to assess the performance of HEIs. Different with other sectors, the evaluation of language learning performance is often conducted through statistical analysis or qualitative methods (Leshem & Bar-Hama, 2008; Richards, 2010; Thaine, 2004). The number of papers applying DEA to language performance is growing, such as McGowan and Graham, 2009, Montoneri et al. (2010a, 2010b, 2011).

DEA can measure the relative efficiency of HEIs from commonly available performance indicators. However, the choice of performance indicators can often affect learning and teaching performance. Therefore, the main purpose of this study is not to decide which indicators are the most suitable, but to design a teaching performance improvement mechanism which can find the priority of the selected evaluated indicators and help to formulate improvement suggestions for educators. This study applies DEA to investigate the indicators contributing to learning and teaching performance and focuses on the indicators coming from students' ratings of teachers (questionnaires filled at the end of each semester) about the course they follow. The results of these ratings are often considered an important reference to evaluate whether teachers' teaching performance can meet one of the criteria demanded by the case school to teachers.

3.1. Teaching performance improvement mechanism

This paper aims at designing a teaching performance improvement mechanism as illustrated in Fig. 1. Each year, the students who enter university have different background and characteristics. Even though freshmen follow the same training, they meet different teachers providing different teaching efforts. As students have different learning efforts, the learning results will be unpredictable. The 4 following categories, teachers, students, input indicators, and output indicators, vary directly each semester. They can be classified as the major dynamic items in the teaching performance improvement mechanism. The objective of this mechanism is to identify the important indicators having a significant impact on learning and teaching performance based on students' ratings of teachers in order to offer suggestions for further teaching improvement.

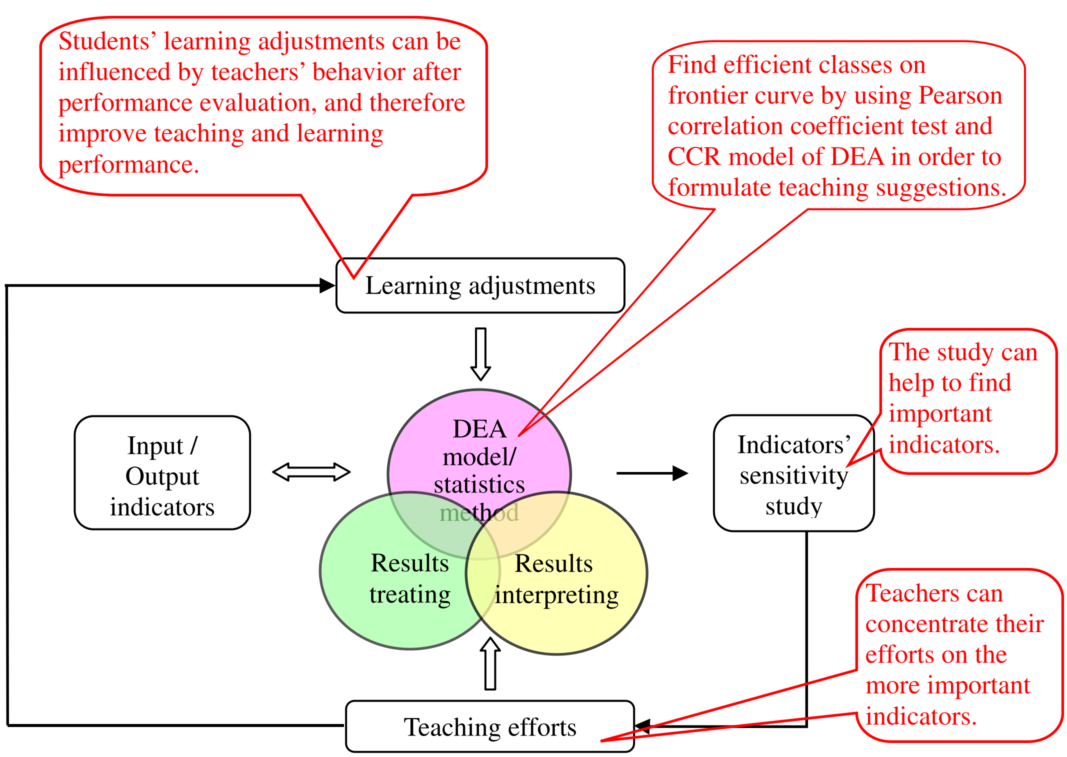

Fig. 1. Diagram of teaching performance improvement mechanism.

In Fig. 1, part of the mechanism is composed of the DEA model and the statistics method, the results treating, and the results interpreting. It can indicate whether the selected indicators have a significant impact on the learning results and can be the representative KPIs or not. If not, the input or the output indicators should be replaced. This evaluating process is presented by the sign with a double sided arrow between the DEA model/Pearson correlation coefficient test and the indicators. Once the indicators are chosen, a sensitivity study will be conducted in order to help teachers to concentrate their efforts on the more important indicators.

3.2. DEA model

DEA is a non-parametric technique used for benchmarking and to measure performance. It can determine the relative efficiencies of a set of homogeneous and comparable units. These entities are commonly called decision making units (DMUs), that is, the evaluated units, which receive multiple inputs and produce multiple outputs (Lee, 2009; Lin et al., 2009). Decision making unit (DMU) was the name used by Charnes, Cooper, and Rhodes (1978) to describe the units being analyzed in DEA. The purpose of the DEA is to establish the relative efficiency of each DMU within a sample (Samoilenko & Osei-Bryson, 2008).

Charnes, Cooper, and Rhodes (1978) expand Farrell's (1957) efficiency measurement concept of multiple inputs and single output to the concept of multiple inputs and multiple outputs converted to single virtual input and output by a linear combination. They estimate the efficiency frontier by the ratio of two linear combinations and measure the relative efficiency of each DMU in constant returns to scale (CRS). As there is a linear relationship between inputs and outputs, an increase in a unit's inputs leads to a proportionate increase in its outputs. This method is now so called "Charnes-Cooper-Rhodes (CCR) model or CCR model". The efficiency value of the CCR model is the overall efficiency of the DMU. DEA is appealing to scholars since it can assess the efficiency of decision making units (DMUs) with multiple inputs and multiple outputs using only information on input and output quantities (Fu & Huang, 2009).

This paper adopts the evaluating method—DEA to build up the learning performance mechanism and perform the efficiency evaluations of a specific course. We investigate the operation performance of DMUs by analyzing these input items and by interpreting the output items according to different domain knowledge. Thus, the DMUs will be the classes that have this course; the input items can be the background or the learning efforts of students, etc.; the output items can be the results of learning such as score, satisfaction, and so on.

Through the analysis of a specific evaluating method, we can better understand the operational efficiency of DMUs and provide the references of concrete and practical strategies for those units with lower operating efficiency. That is, we can focus on the learning performance of each class through the quantitative analysis and provide some concrete and practical learning or teaching strategies for the classes with lower average scores. This notion can be explained by the efficient frontier curve which analyzes, under certain circumstances, how many efforts are necessary for the output performance to come close to the efficient frontier. In our paper, under the same input resources, we firstly focus on the gap between the actual output performance and the target output performance. That is, we would like to clarify how many improvements are necessary to reach the optimal performance of outputs. It is also possible to realize the minimum input resources necessary by fixing the output resources.

3.3. Data selecting—input and output indicators

3.3.1. Ethical considerations

The university studied in this paper gave the authors the authorization to conduct this research and made the needed data available to us. Teacher and student anonymity and confidentiality are respected and protected. Faculty names were deleted and classes were named from D1 to D18.

3.3.2. The data source

The study case is a private church-sponsored university established in 1956 in Taiwan. Approximately 11,000 undergraduates and a little more than 1000 graduates attend the university. The Department of English Language, Literature and Linguistics is one of the oldest departments in the university; it provides both graduate and undergraduate education (more than 700 students). The data comes from the university's online student rating system, which provides student feedback to professors at the end of each semester. Students are required to fill out the feedback.

The characteristics of the research object are as follows:

1. Freshmen students of a department of English in a university of Taiwan, entering from the academic year 2004, 2005, and 2006.
2. They follow the same training program of English conversation for two consecutive semesters to meet the homogeneity of the evaluated object. English conversation is a required course for freshmen.
3. The English conversation course for freshmen is required (2 h/week). Each teacher can choose the text-book of his/her choice. Most of the teachers propose group discussions and role plays during the class. The teachers set various goals, such as: to foster self-confidence in students' ability to spontaneously express their feelings, ideas, and opinions; to enhance students' listening and speaking abilities using various themes; to improve students' abilities for passing TOEFL or GEPT Listening & Speaking.
4. A total of 18 classes taught by full-time teachers (part-time teachers are not included in this paper in order for teachers' characteristics to be more consistent) are selected as the decision making units (DMUs), that is, the evaluated units. They are named from D1 to D18.
5. Each class contains around 50 students.

The characteristics of the data source are as follows:

1. The data are based on questionnaires (10 questions) filled out by the students at the end of each semester for each class. Each question is rated from 1 (very unsatisfied) to 5 (very satisfied) by the students.
2. Beginning with 2007, the university has made some modifications in the questionnaires. Therefore, this paper, for the sake of consistency and because of insufficient number of evaluated classes, selects data prior to 2007. In any case, this paper aims at providing a method to identify the indicators contributing to learning and teaching performance; this method can be applicable to different kinds of data and various types of courses.
3. To ensure the reliability of the questionnaires, at least half of the class must answer seriously. If a student gives ratings too different from the rest of the class, he/she is excluded.
4. The average scores of each question undergo a correlation analysis to test the reliability of the ratings and to find representative indicators in this study.
5. The data concerning the selected indicators is fed in the software Frontier Analyst to calculate the performance values of each evaluated class.

The specification of the outputs and inputs is a crucial first step in DEA. According to Martin (2006), the reliability of the results depends on the accurate selection of the indicators best adapted to the objective of the study. After the rule of thumb, the number of evaluated units is suggested to be two times or even four times the number of indicators. Therefore, based on the questionnaires, four indicators are appropriate in the current study. A correlation coefficient test is conducted in order to understand whether the principle of isotonicity is satisfied and the degree of the correlation between the output items and input items. The input and output indicators chosen for the evaluation model, which are highly correlated, are presented as follows:

Input indicators:

- I1. The richness of course contents: it refers to the degree of teachers' professional knowledge for the preparation of teaching materials.
- I2. The diversity of accessed multiple teaching channels: it indicates whether teachers can increase students' learning interest and learning motivation.

Output indicators:

- O1. The positive degree of teaching attitude: it signifies whether teachers can positively respond to students' questions and the maturity of teachers' teaching skills and communication skills.
- O2. Students' learning performance: it indicates students' learning performance after receiving a period of language training.

All the data acquired are fed into the teaching performance improvement mechanism designed for this paper in order to obtain the relative efficiency values, indicators' priority and improvement suggestions for each inefficient evaluated class.

3.4. Correlation analysis of input and output indicators

Generally speaking, the correlation of the input items and output items in the evaluated units of DEA is commonly verified by the statistics method such as regression analysis, factor analysis, and correlation coefficient test. The correlation analysis used in this paper is the Pearson correlation coefficient test. The higher the Pearson correlation coefficient is, the more closely the relationship between two variables will be; on the contrary, the lower the correlation coefficient is, the lower the correlation between two variables will be. In general, the Pearson correlation coefficient of 0.8 or above represents a very high correlation; the value of 0.6–0.8 represents a high correlation; the value of 0.2–0.4 represents a low correlation; the value inferior to 0.2 represents the extremely low correlation or not correlated. The input and output items listed in Table 4 are abbreviated by I1, I2 and O1, O2 respectively. The correlation coefficients among these 4 indicators are all above 0.8 with a significant level of 1%. This shows a very high degree of correlation. The principle of isotonicity is satisfied.

Table 4
Correlation coefficients between input and output items.

Outputs	Inputs	
	I1 (Richness of teaching materials)	I2 (Diversity of learning channels)
O1 (Positive degree of teaching attitude)	0.974***	0.802***
O2 (Students' learning performance)	0.903***	0.878***

Note: *** denotes significant levels at 1%.

4. Empirical results and suggestions

The calculation is performed in two phases. In phase 1, all the 4 evaluated indicators chosen in Section 3.4 are used in the DEA model. The results of numerical analysis in phase 1 are used to clarify whether the existing teaching methods can achieve the desired results and what are the improved methods. The research results acquired by applying DEA are expected to indicate whether the existing teaching scale is in optimal size and whether students' learning efforts and teachers' teaching efforts will reach the expected performance. In phase 2, a sensitivity study is performed by withdrawing one of the 4 indicators respectively in order to realize the indicators' influence scale on the evaluated DMUs. As a result, we can identify the important and adequate items suitable for the performance evaluation. This paper can provide suggestions for teachers about how to make better use of limited teaching and learning resources. The teaching performance improvement mechanism can be also applied to other fields or other languages in future studies.

4.1. Efficiency analysis of learning performance

The relative efficiency value of the CCR model is the overall efficiency of the evaluated unit. If the efficiency value equals to 1, the evaluated unit is efficient; if the efficiency value is less than 1, the evaluated unit is inefficient (Lee, 2009; Lin et al., 2009). This paper analyzes the learning performance by using the frontier analysis. The CCR score listed in Table 5 refers to the overall efficiency in the CCR model for 18 DMUs, named from D1 to D18. The average efficiency is 0.986. The overall efficiency of the DMUs D1, D4, D6, D10, and D13 shows the best performance with the value of 1. That is, their CCR efficiency value is all on the frontier curve without the need of any further improvement in the inputs and outputs from the perspective of efficiency.

In Table 5, the column "Room for improvement" reveals how many improvements are necessary for the DMU and in what dimension. Since this empirical result is output oriented, we emphasize firstly on how much the insufficiency of output performance is under the current input resources; that is, without any additional input effort. That explains why the values of input dimensions, I1 and I2, are always 0 or negative. For example, the class D3's overall efficiency is the lowest. There is still 4.6% of effort to do in the positive degree of teaching attitude and students' learning performance. Teachers are suggested to respond to students' questions more positively with more detailed explanations and to improve the teaching skills and communication skills so as to meet students' needs. As for the case of the class D9, the improvement value of I2 is −0.5%. Teachers are suggested to lightly reduce the quantity of course contents. As a result, students can better assimilate the basic and important information of the course and enhance their learning performance.

The column "Refs" indicates the number of times the other DMUs are referring to it but not including themselves. For example, there are 12 DMUs referring to the class D4. No DMUs will refer to the inefficient DMUs; 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, that is, the number of efficient units in each inefficient DMU's reference set. For example, the class D3 refers 3 times to other DMUs; that is, D3's reference set contains 3 efficient DMUs which are D4, D10, and D13. D3 is suggested to refer to D4, D10, and D13's inputs and outputs' scales and to seek advice and ideas on how to make their class more attractive.

4.2. Input/output contributions to CCR efficiency

The input/output contribution percentage gives information about the emphasis for each input/output indicator. As a result, it is

Table 5
Overall technical efficiency and relative performance index of evaluated units.

DMU name	CCR score	Room for improvement (%)				Refs	Peers	Contribution (%)			
		O1	O2	I1	I2			O1	O2	I1	I2
D1	1.000	0	0	0	0	9	0	0	100	100	0
D2	0.982	1.9	1.9	0	0	0	3	81.2	18.8	80.9	19.1
D3	0.956	4.6	4.6	0	0	0	3	81.1	18.9	81.4	18.6
D4	1.000	0	0	0	0	12	0	100	0	100	0
D5	0.989	2.4	1.1	0	0	0	1	0	100	100	0
D6	1.000	1.2	0	0	0	0	1	0	100	100	0
D7	0.995	0.5	0.5	0	0	0	3	72.9	27.1	82.9	17.1
D8	0.959	4.3	6.5	0	0	0	2	100	0	81	19
D9	0.982	1.8	1.8	0	−0.5	0	2	67.2	32.8	100	0
D10	1.000	0	0	0	0	8	0	0	100	74.7	25.3
D11	0.972	2.9	11.9	0	0	0	2	100	0	82.7	17.3
D12	0.972	2.9	2.9	0	0	0	3	72.6	27.4	83.6	16.4
D13	1.000	0	0	0	0	7	0	0	100	0	100
D14	0.997	0.3	5.5	0	0	0	2	100	0	80.7	19.3
D15	0.972	2.9	2.9	0	0	0	3	72.9	27.1	83.1	16.9
D16	0.986	1.5	2.8	0	0	0	2	100	0	81.7	18.3
D17	0.997	1.7	0.3	0	0	0	2	0	100	73.6	26.4
D18	0.998	0.2	0.2	0	0	0	3	72.9	27.1	83.1	16.9
Average	0.986	1.62	2.38					56.7	43.3	81.6	18.4

Note: O1 refers to the positive degree of teaching attitude; O2 refers to students' learning performance; I1 refers to the richness of course contents; I2 refers to the diversity of accessed multiple teaching channels.

possible to identify which inputs and outputs have been used or not in determining efficiency. The values listed in Table 5 indicate a percentage of the overall input and output contributions.

For the DMU D1, the contribution values of O1 and O2 in the calculation of the CCR score (overall efficiency) are 0% and 100% respectively and the contribution values of I1 and I2 are 100% and 0% respectively. This means that students' learning performance (O2) of the class D1 can totally explain its overall efficiency, which is only influenced by the input item I1, that is, the richness of course contents. The input and output indicators' average contribution reveals that O1 and I1 are the major indicators in this evaluation mechanism; the proportions are 56.7% and 81.6%, respectively. That is, generally speaking, the positive degree of teaching attitude is the major output indicator and the richness of course contents is the major input indicator. Therefore, teachers should focus on enhancing the preparation of course contents and adapt to students in order to know whether the amount of information is suitable during the class. Students' motivation will probably decrease if there is too much or not enough information to learn.

4.3. Sensitivity study of evaluated indicators

Boussofiane, Dyson, and Thanassoulis (1991) indicate that an additional input or output item will weaken the discrimination of DMUs' efficiency evaluated by the DEA model. If the evaluated indicators of DEA are excessive, the DMUs will all become efficient. Therefore, it will be difficult to observe their performances. By inference, decreasing the number of input and output items will decrease the overall efficiency of evaluated units. Therefore, a sensitivity study by withdrawing an input or output item can help to further identify the characteristics and performance among the efficient DMUs in phase 1. In addition, the result of the sensitivity study can also help to clarify the influences of inputs and outputs on the different evaluated units. The base case is the previous study in phase 1 with 4 indicators (2 inputs and 2 outputs in this paper). The sensitivity study consists here in withdrawing one of the 4 indicators in order to observe its impact on each DMU's CCR score. For example, in case I2–O1–O2, I1 is not included in the study. The variation scale of CCR score compared with the base case can show that the higher the variation scale is, the more sensitive the DMU on the withdrawn indicator will be.

Table 6 lists the overall efficiency and ranking order of evaluated units according to different indicator combinations. The empirical results show that the number of efficient DMUs decreases from 5 for the base case to 1, 3, 4, or 2 for the cases withdrawing indicators I1, I2, O1, or O2, respectively. In addition, the variation percentages of CCR (denoted as ΔCCR) are negative or equal to 0; the average CCR efficiency decreases from the base case's 0.986 to even 0.879 in the case I2–O1–O2 (withdrawing evaluated indicator I1). This confirms the above-mentioned assumption that "withdrawing an input or output item will decrease the overall efficiency of evaluated units".

If I1 is withdrawn (case I2–O1–O2), only one DMU (D13) remains efficient compared with the base case; D4's ΔCCR equal to −16.2% is the maximum value among all the ΔCCR values. The efficient DMU D4 becomes inefficient with the ranking order of 17. This means that the evaluated indicator I1 has the maximum impact on D4 and minimum impact on D13 among all the DMUs. Concerning the average ΔCCR of all the 18 DMUs, the value of case I2–O1–O2 is up to −10.9% which is largely higher than those (varying from −0.6% to −2.0%) of cases I1–O1–O2, I1–I2–O2, and I1–I2–O1.

Similarly, the evaluated indicator I2 has the maximum impact on D13 (ΔCCR is up to −3.8%) and minimum impact on D1, D4, D5, D6 and D9 whose CCR values remain constant compared with the base case. The indicator O1 has the maximum impact on D11 (ΔCCR is up to −8.6%) and minimum impact on D1, D5, D6, D10, D13 and D17. The indicator O2 has the maximum impact on D5 (ΔCCR is up to −2.4%) and minimum impact on D4, D8, D11, D13, D14 and D16.

The sensitivity degree indicated in the Fig. 2 is defined as the variation of DMUs' CCR compared with that of the base case, denoted as the ΔCCR. The negative value of ΔCCR means that the indicator's absence will reduce DMUs' efficiency value. As illustrated in Fig. 2, the sensitivity degree of evaluated indicators on DMUs can clearly indicates that D13 is always efficient for all the cases except the case without taking into account the indicator I2, the case I1–O1–O2. This phenomenon confirms that D13 is more sensitive with the indicator I2. Compared with the indicators I2, O1, and O2, almost all the DMUs are more sensible with I1 except the

Table 6
Comparison of CCR efficiency and ranking according to different indicators combinations.

DMU	Base case		Case I2–O1–O2[a]			Case I1–O1–O2			Case I1–I2–O2			Case I1–I2–O1		
	CCR	Rank	CCR	ΔCCR[b]	Rank	CCR	ΔCCR	Rank	CCR	ΔCCR	Rank	CCR	ΔCCR	Rank
D1	1.000	1	0.869	−13.1%	8	1.000	0.0%	1	1.000	0.0%	1	0.990	−1.1%	7
D2	0.982	13	0.855	−12.9%	14	0.975	−0.7%	11	0.964	−1.8%	12	0.980	−0.2%	9
D3	0.956	18	0.859	−10.2%	12	0.945	−1.2%	17	0.945	−1.2%	15	0.954	−0.3%	18
D4	1.000	1	0.838	−16.2%	17	1.000	0.0%	1	0.973	−2.7%	9	1.000	0.0%	1
D5	0.989	10	0.860	−13.1%	11	0.989	0.0%	6	0.989	0.0%	6	0.966	−2.4%	16
D6	1.000	1	0.869	−13.1%	8	1.000	0.0%	1	1.000	0.0%	1	0.977	−2.3%	10
D7	0.995	9	0.857	−13.9%	13	0.992	−0.3%	5	0.978	−1.7%	8	0.992	−0.3%	6
D8	0.959	17	0.850	−11.3%	15	0.946	−1.4%	16	0.917	−4.4%	17	0.959	0.0%	17
D9	0.982	12	0.830	−15.5%	18	0.982	0.0%	10	0.967	−1.5%	11	0.977	−0.5%	11
D10	1.000	1	0.930	−7.1%	3	0.984	−1.6%	9	1.000	0.0%	1	0.995	−0.5%	4
D11	0.972	14	0.946	−2.7%	2	0.939	−3.4%	18	0.888	−8.6%	18	0.972	0.0%	13
D12	0.972	14	0.879	−9.6%	6	0.962	−1.0%	14	0.970	−0.2%	10	0.966	−0.6%	15
D13	1.000	1	1.000	0.0%	1	0.962	−3.8%	15	1.000	0.0%	1	1.000	0.0%	1
D14	0.997	8	0.870	−12.7%	7	0.987	−1.0%	8	0.924	−7.3%	16	0.997	0.0%	3
D15	0.972	14	0.848	−12.7%	16	0.966	−0.6%	12	0.958	−1.5%	13	0.969	−0.3%	14
D16	0.986	11	0.908	−7.9%	4	0.964	−2.2%	13	0.954	−3.2%	14	0.986	0.0%	8
D17	0.997	7	0.888	−10.9%	5	0.988	−0.9%	7	0.997	0.0%	5	0.975	−2.2%	12
D18	0.998	6	0.869	−12.9%	8	0.992	−0.5%	4	0.983	−1.5%	7	0.994	−0.3%	5
CCR Avg.	0.986		0.879	−10.9%		0.976	−1.0%		0.967	−2.0%		0.980	−0.6%	

[a] "I2–O1–O2" means the studied case takes into consideration only the input I2 and outputs O1 and O2.
[b] ΔCCR means the variation percentage of studied case's CCR score compared with that of the base case.

DMU D11. Generally speaking, O1 is the secondly most influential among the 4 indicators.

In conclusion, the results of Table 6 and Fig. 2 show that the indicator priority is I1 > O1 > I2 > O2; that is, the richness of course contents > the positive degree of teaching attitude > the diversity of accessed multiple teaching channels > students' learning performance. The preparation of course contents and a positive attitude, that is, a good atmosphere during the class, are very important indicators. According to personal observations, Taiwanese students are quite shy and do not like to express their opinions and speak a foreign language in public, especially in front of their classmates (fear to lose face generally associated with a lack of self-confidence). Teachers are suggested to encourage them to practice during the class. They should find a balance between correcting students' mistakes and motivating them so that they are not afraid to participate. As to students, they have to show their efforts during the training and accept criticism. However, if they are over criticized by their teachers, they will probably lose their motivation. It is always preferable to tell students in private what their problems are (pronunciation, speaking louder, articulation, grammatical mistakes, too poor contents and vocabulary, etc…). It is important they do not feel humiliated in public.

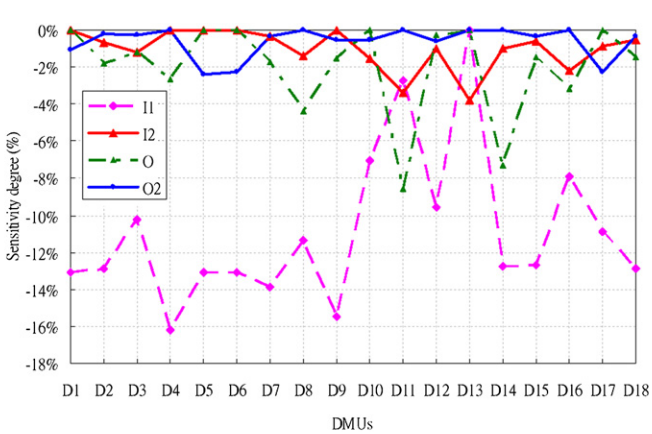

Fig. 2. Sensitivity degrees of evaluated indicators on DMUs.

4.4. Case study: improvement suggestions for an inefficient DMU

We provide a detailed demonstration of indicators' importance by studying the case of an inefficient DMU, D3. The analysis procedure is divided into two parts: base case study and indicator's sensitivity study, in order to formulate improvement suggestions for inefficient DMUs. The relative data are presented in Table 7.

4.4.1. Part 1: base case study

It consists in taking into consideration all the input/output indicators in calculating all the DMUs' relative efficiency values by applying DEA method in order to identity the inefficient and efficient DMUs. Then, we list each inefficient DMU's reference set members' contributions to the inputs'/outputs' optimal values. These contributions are ranked item by item in order to know efficient DMUs' impact order. Considering input/output items' room for improvement and their contribution in calculating the relative efficiency helps to provide some clues in finding indicators' importance.

4.4.2. Part 2: indicator's sensitivity study

The indicator's sensitivity study of all the DMUs consists in withdrawing indicators one after another in order to observe the impact of indicator's absence on DMUs' efficiency values. We can obtain each indicator's sensitivity degree and their impact on the ranking variation compared with the base case. The indicator's sensitivity study helps to more objectively formulate improvement suggestions for the inefficient DMU.

4.4.3. Improvement suggestions for case study D3

In Part 1, the base case study, D3's relative efficiency value is 0.956. The efficient DMUs composing D3's reference set are D4, D10 and D13.

1. All the D3's output items should be improved equally to 4.57%.
2. All the D3's inputs, I1 (richness of course contents) and I2 (diversity of teaching channels), can be maintained at the same level.
3. All the input and output items have a contribution in calculating D3's relative efficiency, notably O1 (positive degree of

Table 7
Efficiency improvement analysis for the inefficient DMU D3.

	Outputs		Inputs	
	O1	O2	I1	I2
Part 1	*Base case study*			
Reference set's contributions (%) to indicators' benchmark values				
D4	36.3 (2)	35.7 (2)	35.7 (2)	38.4 (2)
D10	56.0 (1)	56.7 (1)	56.4 (1)	54.8 (1)
D13	7.7 (3)	7.6 (3)	7.9 (3)	6.8 (3)
Room for improvement (%)	4.57	4.57	0	0
Outputs/Inputs contribution (%) in calculating relative efficiency	81.1	18.9	81.4	18.6
Part 2	*Indicator's sensitivity study*			
Sensitivity degree (%)	−1.2	−0.3	−10.2	−1.2
Ranking variation in sensitivity study compared with the base case	+3	0	+6	+1

Note: 1. O1 refers to the positive degree of teaching attitude; O2 refers to students' learning performance; I1 refers to the richness of course contents; I2 refers to the diversity of accessed multiple teaching channels. 2. The numbers in the parentheses indicate the contributions ranking for each input/output item.

teaching attitude representing 81.1%) and I1 (richness of course contents representing 81.1%).

4. Taking into account the I/O items' room for improvement and contribution in calculating efficiency, the items with values not equal to zero at the same time should be improved in a priority in order to rapidly increase the DMU's relative efficiency. Therefore, D3 can only make efforts on O1 and O2, notably on O1 which represents 81.1%.
5. If D3 hopes to rapidly increase its relative efficiency, D3 is suggested to mainly refer to D10's positive degree of teaching attitude (O1) up to 56.0%, to D4's O1 to 36.3% and to D13's O1 to 7.7%; then refer to D10's students' learning performance (O2) up to 56.7%, to D4's O2 to 35.7% and to D13's O2 to 7.6%.
6. If D3 hopes to increase its overall performance in each I/O item in the long term, its performance improvement measures can not merely refer to one single efficient DMU, even though D10 is the major model for D3. D3 is suggested to mainly refer to all the D10's input and output items around 56%, then refer to all the D4's items around 36%, and finally refer to all the D13's items around 8%.

In Part 2, the indicator's sensitivity study, D3's efficiency values are 0.859, 0.945, 0.945, and 0.954 in an indicators' sensitivity study by withdrawing I1, I2, O1, and O2, respectively.

7. The indicator's sensitivity study shows that I1 (richness of course contents) has the major impact on D3's relative efficiency; the withdrawal of I1 can reduce D3's relative efficiency by up to 10.2%, from 0.956 to 0.859.
8. Even though O1 and I2 have the same sensitivity degree (−1.2%), the withdrawal of O1 helps D3's ranking among all the DMUs gaining 3 places instead of 1 place in the case of withdrawing I2. Therefore, O1 (the positive degree of teaching attitude) is slightly more important than I2 (diversity of teaching channels) for D3.
9. We conclude that for the case study D3, the indicators' priority is I1 > O1 > I2 > O2.

4.5. A brief qualitative study of factors affecting teaching and learning performance

In addition to above-mentioned evaluated indicators, various factors can also have an impact on inefficient DMUs. A brief qualitative exploration of some factors affecting teaching and learning performance is conducted after a few informal interviews, discussions and even some personal observations. Some factors affecting teachers and students are presented as follows:

For teachers:

1. Teacher seniority: according to students' ratings at the end of each semester, experienced teachers do not necessarily obtain higher scores. It could probably be because some of them are not as enthusiastic as the new teachers and because they do not update their teaching contents enough (high-tech channel, online courses). According to Centra (1993), the greatest teaching improvement occurs during a teacher's first few years of teaching.
2. Teaching load: teachers teaching many different courses or changing their courses every year may have a lower score because they spend less time preparing for each course and because they do not have time to accumulate enough knowledge and experience.
3. Research: teachers who can guide students in their area of expertise may have a better evaluation, if they are able to adapt to the level of their students. However, they should not lower the level in order to obtain a better evaluation from the students. Wang (2010) shows that only 60.3% teachers combined teaching and researching to improve teaching quality.
4. Administration load: some teachers have a lot of administration obligations (meetings, promotion of the university, organization of various activities, such as the English festival, cooperation with enterprises, being tutors, writing recommendation letters…). These activities are time-consuming and may affect not only teaching, but also research.
5. Teachers' information skill: teachers who can design multimedia courses and use Internet to communicate with students can improve and enrich the teaching contents and skills. According to Carlson and Gadio (2002), one of key determining factors for improving student performance is the use and application of technology by teachers, such as learning from the Internet. In Taiwan, Chen (2008) studies the factors influencing teachers in integrating the Internet into their instruction and shows that continuous professional development focused on technology application in language instruction is imperative.

For students:

1. Part-time work: some students are able to find a job in relation with English, such as being a teacher in a cram school, working for a publishing company, or being a translator during international events. In this case, having a work may be beneficial and give students opportunities to practice English and meet

foreigners. If the part-time job has nothing to do with English, it may distract students from their courses.

2. Professional license: in Asia, students are encouraged to take English proficiency tests, which can help when applying for a grant, entering a graduate school, and eventually finding a job. For example, the TOEIC (Test of English for International Communication) test and the TOEFL (Test of English as a Foreign Language) can measure the English skills of foreign students and entice them to improve their level.
3. Participation in associations and clubs: even though these activities can enrich students' life and allow them to meet people with similar interests, they are time-consuming and exhausting. Students who participate in too many activities reduce the amount of time left for preparing their courses and exams accordingly. In order to increase their English learning performance, students are suggested to participate in one or two associations in relation with English (theater club, English Corner…).
4. Personality: some students are more energetic and enjoy participating in the class. But most students have difficulty to express their feelings and their ideas in a foreign language. The teacher should try to motivate students to participate during the class and give their opinion. Taylor et al. (2004) design a four-cell matrix of teaching practice classification and showed that the teaching and learning performances are improved when the teacher is the facilitator and when the teaching is student-centered.
5. Learn other foreign languages than English: for students speaking non-Latin languages, English is relatively more difficult to learn. It is indeed easier for French students to learn English and vice versa as around 70% of the vocabulary is identical in the two languages. Meyerhoff (2009) adapt a study to Japanese students by Yuan, Liberman and Cieri (2006) which showed that fluent speakers of ESL from various nationalities have varying rates of speaking in English. He stated that fluent Japanese English learners spoke English the slowest, and fluent French English speakers spoke English most rapidly.
6. Teamwork: Clinton and Kohlmeyer (2005) show that students doing teamwork during the class express greater motivation and rated their teacher higher even if they did not show evidence of any significantly different performance results.

5. Conclusions

5.1. Findings

This paper applies data envelopment analysis (DEA) to explore the relative overall efficiency of evaluated units and to identify which are the efficient ones. The overall efficiency of D1, D4, D6, D10, and D13 shows the best performance with the value of 1. Inefficient DMUs can know how much effort they need to do by looking at the performance indicators' room for improvement; they can also seek advice and ideas on how to make their class more attractive in order to improve their performance from the efficient DMUs in their own reference set.

The sensitivity study performed in this paper can also help to clarify the influences of inputs and outputs on the different evaluated units. The results show that the number of efficient DMUs decreases from 5 for the base case to 1, 3, 4, or 2 for the cases withdrawing indicators I1, I2, O1, or O2, respectively. In addition, the average relative efficiency decreases from the base case's 0.986 to even 0.879 in the case I2–O1–O2 (withdrawing evaluated indicator I1). This confirms the assumption that withdrawing an input or output item will decrease the overall efficiency of evaluated units. It appears that the evaluated indicator I1 has the maximum impact on D4; I2 has the maximum impact on D13; O1 has the maximum impact on D11; and O2 has the maximum impact on D5. A detailed analysis of the data shows that the indicator priority is I1 > O1 > I2 > O2; that is, the richness of course contents > the positive degree of teaching attitude > the diversity of accessed multiple teaching channels > students' learning performance.

This paper presents then the case study of one inefficient DMU, D3. This DMU has a relative efficiency value is 0.956. I1 (richness of course contents) has the major impact on D3's relative efficiency. D3 is suggested to mainly refer to all of D10's input and output items around 56%, then to refer to all of D4's items around 36%, and finally to refer to all of D13's items around 8%. Concerning D3, the indicators' priority is, like for most DMUs, I1 > O1 > I2 > O2.

Finally, a brief qualitative study shows that some of the factors affecting teachers' performance are teacher seniority, teaching, research, and administration loading, and teachers' information skills. Concerning the problem of burnout (emotional exhaustion and depersonalization), it seems that it affects teaching more than research productivity. Some of the factors affecting students' performance are part-time work (very few students have a full-time job), proficiency tests, participation in associations and clubs, personality, learning a second or third foreign language, teamwork, and health.

5.2. Implications

Various studies select different evaluated indicators according to their availability, to the purpose of their research, and to each country's culture of performance. The main contribution of this study is not to decide which indicators are the most suitable, but to design a teaching performance improvement mechanism which can find the priority of the selected evaluated indicators and help to formulate improvement suggestions for educators. Therefore, our demonstration on how to screen primary indicators will be quite useful for further studies in other countries or fields. The results of this paper can serve as a model for decision-makers to design the educational policies satisfying the objectives of enhancing the competitiveness of HEIs as well as the goal of the Ministry of Education.

5.3. Limitations and future directions

The results of the study have to be interpreted in light of its limitations. DEA only gives efficiencies relative to the data considered. Some DMUs may be efficient simply because no one else is competing in that range (Johnes & Johnes, 1993). Moreover, it might be possible to improve the performance of even the efficient units (Martin, 2006). Our previous study also demonstrates that the efficient evaluated classes do not necessarily have higher actual values of inputs and outputs than the inefficient ones, but in other ranges (Montoneri et al., 2011). However, the DEA model provides few clues on how to improve the performance of empirically efficient units (Sowlati & Paradi, 2004). This study offers suggestions to teachers on how to improve their teaching and proposes a method to find out on which indicators they should focus as a priority. Future studies could not only analyze teachers' response to student evaluation of teaching, but how they would respond to improvement suggestions. It would also be interesting to explore students' perceptions of the faculty and class evaluation process.

Acknowledgments

The corresponding author would like to thank the National Science Council of the Republic of China (Taiwan) for financially supporting this research under contract no. 99-2914-I-126-017-A1. The primary results of this paper come from a conference paper entitled "Application of DEA on teaching resource inputs and learning performance" presented at TARC International Conference 2010, Kuala Lumpur, Malaysia. The authors would also like to thank the TIC 2010 Editorial Committee and the reviewers who awarded this conference paper the "outstanding paper for sub-theme, Methodologies and Strategies in Learning, Teaching and Assessment". The authors would like to thank Ms. Catharina Lemmer (Foreign Language Center, Providence university) and all the reviewers who contributed to improve this paper.

References

Abbott, M., & Doucouliagos, C. (2003). The efficiency of Australian universities: a data envelopment analysis. *Economics of Education Review, 22*(1), 89–97.

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.

Australian Learning and Teaching Council. (2008). *Review of Australian and international performance indicators and measures of the quality of teaching and learning in higher education*. Report commissioned by the Department of Education, Science and Training. Camberra, Australia: Commonwealth of Australia.

Avkiran, N. (2001). Investigating technical and scale efficiencies of Australian universities through data envelopment analysis. *Socio-economic Planning Sciences, 35*, 57–80.

Azma, F. (2010). Qualitative indicators for the evaluation of universities performance. *Procedia Social and Behavioral Sciences, 2*(2), 5408–5411.

Bologna Declaration. (1999). *Joint declaration of the European Ministers of education. The European Higher Education Area*. Retrieved June 20, 2011 from. http://www.bologna-berlin2003.de/pdf/bologna_declaration.pdf.

Boussofiane, A., Dyson, R. G., & Thanassoulis, E. (1991). Applied data envelopment analysis. *European Journal of Operational Research, 52*, 1–15.

Burden, P. (2008). Does the use of end of semester evaluation forms represent teachers' views of teaching in a tertiary education context in Japan? *Teaching and Teacher Education, 24*(6), 1463–1475.

Carlson, S., & Gadio, C. T. (2002). Teacher professional development in the use of technology. In J. Sikula (Ed.), *Handbook on teacher education*. New York: Macmillan.

Centra, J. A. (1993). *Reflective faculty evaluation. Enhancing teaching and determining faculty effectiveness*. San Francisco, USA: Jossey-Bass.

Chalmers, D. (2008). Teaching and learning quality indicators in Australian universities. OECD/IMHE Conference: Outcomes of higher education: Quality relevance and impact. 8–10 September 2008, Paris, France. Retrieved June 20, 2011, from: http://www.oecd.org/dataoecd/4/40/41216416.pdf.

Chalmers, D., & Thomson, K. (2008). *Snapshot of teaching and learning practice in Australian Higher Education Institutions*. Sydney, NSW: Carrick Institute for Learning and Teaching in Higher Education Ltd.

Charnes, A., Cooper, W. W., Lewin, A. Y., & Seiford, L. M. (1994). *Data envelopment analysis*. Boston: Kluwer.

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

Chen, Y. L. (2008). A mixed-method study of EFL teachers' Internet use in language instruction. *Teaching and Teacher Education, 24*(4), 1015–1028.

Cherchye, L., & Post, T. (2003). Methodological advances in DEA: a survey and an application for the Dutch electricity sector. *Statistica Neerlandica, 57*(4), 410–438.

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*, 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.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society Series A. General, 120*(3), 253–282.

Ferrier, G., & Knox Lovell, C. A. (1990). Measuring cost efficiency in banking: econometric and linear programming evidence. *Journal of Econometrics, 46*(1–2), 229–245.

Førsund, F. R., & Sarafoglou, N. (2002). On the origins of data envelopment analysis. *Journal of Productivity Analysis, 17*(1–2), 23–40.

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* (pp. 197–215). Physica-Verlag, HD, Contributions to Economics.

García-Aracil, A. (2006). Overview of regional differences in the Spanish higher education system. In A. Bonaccorsi, & D. Cinzia (Eds.), *Universities and strategic knowledge creation*. UK: Edward Elgar Publishing Ltd.

García-Aracil, A., & Palomares-Montero, D. (2008). *Changes in universities' efficiency over the time: Differentials according to the missions*. Barcelona: First ISA Forum of Sociology: Sociological Research and Public Debate. Retrieved June 20, 2011 from. http://digital.csic.es/handle/10261/10097.

García Valderrama, T. (1996). *La Medida y el Control de la Eficiencia en las Instituciones Universitarias*. Valencia: Sindicature de Comptes de la Generalitat Valenciana.

Glass, J. C., McCallion, G., McKillop, D. G., Rasaratnam, S., & Stringer, K. S. (2006). Implications of variant efficiency measures for policy evaluations in UK higher education. *Socio-Economic Planning Sciences, 40*(2), 119–142.

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*, 153–175.

Jemric, I., & Vujcic, B. (2002). Efficiency of banks in Croatia: a DEA approach. *Comparative Economic Studies, 44*(2), 169–193.

Johnes, J. (2006). Measuring teaching efficiency in higher education: an application of data envelopment analysis to economics graduates for UK universities. *European Journal of Operational Research, 174*, 443–456.

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

Johnes, J., & Taylor, J. (1990). *Performance indicators in higher education: UK universities*. Birmingham: The Society for Research into Higher Education and Open University Press.

Joumady, O., & Ris, C. (2005). *Determining the relative efficiency of European Higher Education institutions using DEA*. The Netherlands: Maastricht University Research Centre for Education and the Labour Market.

Kozmetsky, G., & Yue, P. (1998). Comparative performance of global semiconductor companies. *Omega (Oxford), 26*(2), 153–175.

Lai, Y. H. (2007). Use of data envelopment analysis to assess the relative efficiency of laptop computer manufacturers: an empirical study. *International Journal of Management, 24*(2), 289–295.

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.

Leshem, S., & Bar-Hama, R. (2008). Evaluating teaching practice. *ELT Journal, 62*(3), 257–265.

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.

Martin, E. (2006). Efficiency and Quality in the Current Higher Education Context in Europe: an application of the data envelopment analysis methodology to performance assessment of departments within the University of Zaragoza. *Quality in Higher Education, 12*(1), 57–79.

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

McMillan, M., & Datta, D. (1998). The relative efficiencies of Canadian universities: a DEA perspective. *Canadian Public Policy, 24*(4), 485–511.

Meyerhoff, S. A. (2009). Analysis of communication strategies used by freshman active English students using YackPack for homework-based speaking tasks. *Asian EFL Journal, 11*(2), 188–213, Retrieved June 20, 2011 from. http://www.asian-efl-journal.com/June_09_sam.php.

Ministry of Education in Taiwan. (2011a). *Important information for education statistics – Statistics for the number of tertiary institutions*. Retrieved June 20, 2011, from. http://www.edu.tw/files/site_content/B0013/overview03.xls.

Ministry of Education in Taiwan. (2011b). *Important information for education statistics – Statistics for the number of students admitted to colleges and acceptance rate*. Retrieved June 20, 2011, from. http://www.edu.tw/files/site_content/B0013/overview09.xls.

Ministry of Interior in Taiwan. (2010). *Sixth weekly report, statistical bulletin 2010*. Retrieved June 20, 2011, from. http://www.moi.gov.tw/stat/news_content.aspx?sn=3829.

Montoneri, B., Lee, C. C., Lin, T. T., & Huang, S. L. (2010a). Application of DEA on English Writing Learning Performance. In Y. C. Chao, & W. C. Wu (Eds.), *Integration and innovation of English language, literature, and linguistics in the EFl classroom. Asian EFL Journal International Conference*, (pp. 588–595). Republic of China (Taiwan): Providence University.

Montoneri, B., Lee, C. C., Lin, T. T., & Huang, S. L. (2010b). Application of DEA on teaching resource inputs and learning performance. In S. N. Chin (Ed.), *Emerging trends in higher education learning and teaching. TARC International Conference on Learning and Teaching*, (pp. 31–38). Malaysia: Tunku Abdul Rahman College.

Montoneri, B., Lee, C. C., Lin, T. T., & Huang, S. L. (2011). A learning performance evaluation with benchmarking concept for English writing courses. *Expert Systems with Applications, 38*(12), 14542–14549.

Ng, Y. C., & Li, S. K. (2000). Measuring the research performance of Chinese higher education institutions: an application of data envelopment analysis. *Education Economics, 8*(2), 139–156.

Pina, V., & Torres, L. (1995). Evaluación del rendimiento de los departamentos de contabilidad de las universidades españolas. *Hacienda Pública Española, 135*(4), 183–190.

Pugh, G., Coates, G., & Adnett, N. (2005). Performance indicators and widening participation in UK higher education. *Higher Education Quarterly, 59*(1), 19–39.

Richards, J. C. (2010). Competence and performance in language teaching. *A Journal of Language Teaching and Research, 41*(2), 101–122.

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.
Sowlati, T., & Paradi, J. C. (2004). Establishing the practical frontier in data envelopment analysis. *Omega – International Journal of Management Science, 32*(4), 261–272.
Taylor, C., Jones, A., Shindler, J., & Cadenas, H. (2004). *"Architect and steward: Shaping a vision of learning". Examining the role of the principal in the immersion of new teachers into existing urban school climates*. Los Angeles: Charter College of Education, California State University. Retrieved June 20, 2011, from. http://www.calstatela.edu/centers/schoolclimate/research/architect_steward.html.
Thaine, C. (2004). The assessment of second language teaching. *ELT Journal, 58*(4), 336–345.
Thore, S., Phillips, F., Ruefli, T. W., & Yue, P. (1996). DEA and the management of the product cycle: the U.S. computer industry. *Computers and Operations Research, 23*(4), 341–356.
Tomkins, C., & Green, R. (1988). An experiment in the use of DEA for evaluating the efficiency of UK university departments of accounting. *Financial Accountability and Management, 4*(2), 147–164.
Valdmanis, V. G. (1990). Ownership and technical efficiency of hospitals. *Medical Care, 28*(6), 552–561.
Wang, L. (2010). An investigation of the current state of college teachers' teaching quality and teacher development. *Asian EFL Journal, 12*(1), 261–284.
Wolf, Z. R., Bender, P. J., Beitz, J. M., Wieland, D. M., & Vito, K. O. (2004). Strengths and weaknesses of faculty teaching performance reported by undergraduate and graduate nursing students: a descriptive study. *Journal of Professional Nursing, 20*(2), 118–128.
Wu, Y. (2010). Performance evaluation of higher education based on CORPS. *Higher Education Development and Evaluation, 5*, 54–61.
Wu, Y., & Li, C. (2009). Research on performance evaluation of higher education based on the model of BSC-DRF-DEA. 16th International Conference on Industrial Engineering and Engineering Management, IE&EM '09, 2030–2034.
Yang, J. (2005). *Quantifying the technical efficiency of Canadian paratransit systems using data envelopment analysis method*. A thesis presented to the University of Waterloo, Belgium.
Yuan, J., Liberman, M., & Cieri, C. (2006). *Towards an integrated understanding of speaking rate in conversation*. USA: Department of Linguistics, Linguistics Data Consortium, University of Pennsylvania. Retrieved June 20, 2011 from. http://www.ldc.upenn.edu/myl/llog/icslp06_final.pdf.
Zhou, C., & Wang, M. (2009). The Evaluation Research on Higher Education Efficiency with Data Envelopment Analysis (DEA). International Conference on Management and Service Science, MASS '09, Beijing, 1–7.