THE ACCOUNTING EDUCATORS' JOURNAL
Volume XX
2010
pp. 21-32

Course Evaluation Validation using Data Envelopment Analysis

Joseph Sarkis
Clark University

Inshik Seol
Clark University

Abstract

In this paper we detail a methodology and some of its variations with respect to evaluating the validity of a teaching and course evaluation tool. The methodology applies a mathematical programming based multifactor productivity approach called data envelopment analysis (DEA). The paper shows how DEA can be effectively applied to check the validity of the evaluation instrument using actual accounting students' responses for an accounting instructor and course. The results show that DEA can be used as supplements to existing course evaluation systems by making the interpretation of the course evaluation instrument more meaningful.

Introduction

Performance evaluation criteria are major concerns to most, if not all, accounting (as well as other) faculty at higher education institutions. For example, Hunt et al. (2009) found that accounting faculty are quite concerned with the performance evaluation criteria used for promotion and tenure decisions when they search for new jobs and/or relocation. One of the most widely used evaluation criteria is teaching effectiveness. Read et al. (2001) surveyed 250 administrators of accounting programs in AACSB-accredited institutions and found that teaching effectiveness was weighed between 34 to 50 percent in tenure evaluation and promotion decisions. Thus, the importance of effective and accurate teaching evaluation and performance criteria cannot be understated for academic faculty.

Among many different teaching evaluation methods (e.g., self-assessment, peer review, outside consultants, etc.), the most popular method is the evaluation by students. Petersen et al (2008) argue that most institutions of higher education use student evaluation because it provides administrators with useful information for faculty reappointment, tenure, and promotion processes, as well as for merit and teaching awards. Specifically, student teaching evaluation is popular because (1) it provides direct feedback to the faculty, (2) it guides faculty for better pedagogical performance in the classroom (Petersen et al (2008). Other advantages of student evaluation over other evaluative methods include arguments that students are: (1) customers; (2) not biased compared to a one-time outside reviewer; (3) an inexpensive way of collecting data; and (4) anonymous (Seiler & Seiler 2002). Read et al. (2001), in their study, found that course evaluation received more than 50 percent of the relative weight for evaluating teaching. Crumbley & Fliedner (2002) also found that accounting administrators would not replace the current course evaluation system with an alternative evaluation system even though there are some problems with their current student evaluation system.

One of the problems of student course evaluation, however, is its reliability and validity. Even though some papers show that student evaluations are valid and reliable mechanisms for teaching evaluation (Wachtel, 1998), others do not agree (Nimmer & Stone, 1991). The common findings are that student course evaluations are affected by many other factors including student characteristics (e.g., expected grade), instructor characteristics (e.g., gender), and course characteristics (e.g., class size), and other environmental characteristics (e.g., ambience of the classroom).

Crumbley et al. (2001) surveyed over 500 accounting students to determine accounting students' perceptions on teaching evaluation and found that students punish instructors who engage in a number of well-known learning/teaching techniques, which encourages instructors to increase student evaluation scores by sacrificing the learning process. They conclude that using student data as a surrogate for teaching performance is an illusionary performance measurement system and suggest other measures should be employed.

Most studies, however, focus more on outside (e.g., situational) factors and not many papers discuss the validity of the students' responses to the course evaluation or how to assess that validity. This paper tries to fill this gap by introducing a methodology to investigate the validity of a course evaluation instrument. This validity is based on the student evaluation responses on 'input' and 'output' evaluation items from a course evaluation questionnaire and the level of consistency (bias) between these inputs and outputs. The methodology relies on a multi-factor productivity model called Data Envelopment Analysis (DEA). DEA allows for the simultaneous comparison of multiple input and output factors that will determine the relative evaluations of a number of students. We apply DEA to an analysis of the valuation instrument for an accounting instructor. The following section introduces the DEA methodology.

DEA: A SIMPLE GRAPHICAL EXAMPLE[1]

To help understand some of the basic foundations of DEA, we will provide a simple graphical example. This graphical example is easy to show when looking at the "envelopment" side of DEA analysis (as mathematically described in expression (A4) in the appendix).

Let us assume that we are using the teaching evaluations of three students, A, B, and C, within an accounting class. We will assume there is only one input (e.g. Instructor Class Preparation rating) and this evaluation is the same for each student (say a 3 rating). Looking at Figure 1, we see that there are two performance 'output' factors reported by students, represented by ratings for "overall teacher effectiveness," along the Y axis, and "knowledge gained from class rating", along the X axis. Thus, we can see that given an equal initial rating on the input, the higher the number on the Y and X axes, the better the evaluated performance of the instructor (assuming a 0-4 scoring range, with higher numbers representing better evaluations). Figure 1 shows that Student A evaluates the knowledge gained from the course relatively highly (a 4 rating), but gave a poor rating for overall teacher effectiveness, while student C gave a high teacher effectiveness rating (4) and a lower knowledge gained rating (1). Student B gave rankings of 1 for teacher effectiveness and 2 for knowledge gained.

If it can be assumed that linear combinations of the student evaluations are allowed, then the line segment connecting students A and C show the possibilities of outputs for virtual students of varying compositions of the "efficient" student evaluations. Similar line segments can be drawn between A and B or between B and C. However, since the segment AC lies beyond segments AB and BC (towards the upper right side of the quadrant), this means that a combination of only A and C will produce an efficient virtual student evaluation score that will generate the highest rating output for a given set of inputs ratings. This line

[1] This simple graphical example is based on Tim Anderson's baseball example located at http://www.emp.pdx.edu/dea/homedea.html

(segment AC) is called the efficient frontier. We see that Student B's evaluation lies below the efficient frontier and can be considered 'inefficient'. Since Student B is within the efficient frontier, it is deemed inefficient. Student B's relative efficiency score is determined by comparing it to a virtual student evaluation V, formed by combining student evaluation points A and C. The virtual student evaluation V, is a linear combination of approximately 64 percent of student C's evaluation and 36 percent of Student A's evaluation.

The efficiency of student B's evaluation is then calculated by finding the fraction of input that student V's evaluation rating would need to be with the Student B's output evaluation scores. This score is calculated by looking at the line from the origin, O, to V. The efficiency of student B's evaluation score is OB/OV, or about 68 percent. Mathematically, we can expand this evaluation to numerous dimensions for inputs and outputs as shown in the Appendix.

Application of the DEA model: A Case example

To show the applicability of the DEA approach to evaluate the validity of student evaluation instruments, we used a sample size of 99 student responses from four accounting classes for an individual accounting instructor. The evaluation instrument is composed of (1) twelve evaluation questions that use a Likert scale rating scheme ranging from a rating of 1 (poor) to 5 (excellent) and (2) other questions which include student characteristics as well as open-end questions.
We selected three inputs and two outputs to show the robustness of the solution and to not overly complicate the analysis. The three inputs include the student evaluation scores of:

1. Instructor's ability to present material clearly.
2. Instructor's preparation for classes.
3. Instructor's overall organization of the course.

The two outputs include the student evaluation scores of:

1. Contribution of this course to students' acquiring new knowledge.
2. Overall effectiveness of the instructor.

The selection is based on the previous research showing that students consider preparation, classroom presentation, organization among others are important dimensions of their learning (O'Toole et al., 2000; Tang, 1997). These evaluation input and output factors were also selected because most of the students fully responded to them (students with 'not applicable' responses in their evaluations for our selected inputs and outputs were not included in the DEA evaluation) and thus allowed us to have a more complete data set for evaluation.

The raw data and results of efficiency scores for each individual are shown in Table 1. To get a better description of the efficiency score results and analyzing them for validity/bias we have graphed them in Figure 2.

If the results across all students were consistent (e.g. all 3 ratings on inputs, all 3 ratings on outputs), then all the students would have an equal relative efficiency score of 1.00. This result would mean that the graphics would be represented with a straight line across at 1.00. Yet, if any one of the students is inconsistent with lower input evaluation rating and higher output evaluation rating, representing a more "efficient" solution, then that student evaluation would get a 1.00 score and other scores will probably move down in relative score. If it were only one student, the remaining students may still be relatively consistent, and a prevalence of the results would still be around a certain efficiency score or couple of scores, depending on the number of input and output factors. If there is more dispersion and many levels of DEA scores, then we could argue that there is less validity in the rankings.

Note that high efficiency scores do not equate to bias. Bias (or inconsistency) refers to the variance in the efficiency scores, not the actual efficiency scores. Low inputs with high outputs lead to the most efficient scores. In some cases high efficiency scores are outliers. If the rest of the sample consistently applies their individual weighting schemes, the other decision making units will have lower but relatively similar efficiency scores.

Such bias may occur from one of two types of occurrences. The first bias would occur when students give high input ratings and lower output ratings (meaning a smaller relative efficiency DEA score, further away from a value of 1.00). The second bias would occur when students provide lower input ratings and higher output ratings (meaning a higher relative efficiency score, closer to a value of 1.00). Thus, larger dispersions in efficiency scores would represent bias in the instrument.

We have two ways of observing bias with DEA-based data. The first is to look to see if there are relatively consistent efficiency scores by determining the heterogeneity of the data using cluster analysis (or levels of lines on a graph); the second is to investigate the dispersion of the data.

As mentioned, clustering and cluster analysis of the efficiency scores may provide insight into the validity and bias of the student evaluations. When a larger number of clusters, with a non-trivial set of students assigned to these clusters, exists, then it is very likely that there are biases and less validity in the responses. Looking at size and number of clusters and the amount of diffusion in the data together will give a better idea of the overall validity of the student responses based on the input and output factors. These clusters may also be statistically evaluated to determine statistical significance in their differences.

A cluster analysis is carried out on our data set using SPSS software and its Two-Step Cluster analysis module. We used a Euclidean distance measure and two separate clustering criterion (Akaike's Information Criterion and Schwarz's Bayesian Criterion). For dispersion of data we used the simple standard deviation and coefficient of variation statistic.

The results show that there are three clusters and the coefficient of variation was about 0.13. Since perfectly unbiased results will have only one cluster and coefficients of variation equal to zero, we clearly have some bias in the results.

The determination of what is acceptable and not acceptable bias (egregiousness of the bias) will typically be judgmental. Thus, let us take a look at the characteristics of the clusters. Table 2 (a) shows that there is 6 percent, 62 percent, and 32 percent in each respective cluster. The raw data shows that the first cluster of evaluation scores are those where the respondents have undervalued the inputs and/or overvalued the output of the evaluations (e.g. student number 3 who assigned a value of 3 for one of the inputs while assigning a value of 4 for the outputs). This was the smallest cluster with only 6 percent having inconsistent evaluations making the instructor and/or class seem better than expected.

If we look at the elements within the second cluster, we see less bias and more consistency in the data here (e.g. students 2 and 4 who rated “5” for all our factors). This is the largest cluster, which signifies that a large percentage of the students were consistent in their appraisal.

The final cluster is comprised of those students who put higher values on the input elements and lesser value on the output elements. Examples of these are shown by student 1 who ranked the inputs as ‘excellent,’ but the outputs as ‘good’ (ratings of “4” and “5”) respectively.

Given these observations, we do not see the extreme weightings and differences where definite biases do exist, one way or the other. The dispersion of the data and the number of clusters do not necessarily point to great disagreement. Thus, we can state with some confidence that the instruments for this sample of classes did not contain much bias. Table 2 summarizes the results.

SUMMARY AND DISCUSSION

The course evaluation by students is the most widely used tool to measure teaching effectiveness of college instructors. However, many previous researchers have questioned the validity of this evaluation approach and its instruments. Our paper introduces a multi-factor productivity model called Data Envelopment Analysis (DEA) and shows how DEA can be applied in checking the validity of course evaluation instruments. Using actual data we show how the results can be interpreted and further evaluated using standard statistical tools.

DEA, however, has some limitations too. For example, DEA techniques tend to use extreme weightings to make a student relative efficiency score as large as possible. To overcome this limitations weight restrictions or range of weight restrictions (assurance regions) may be introduced by decision makers and analysts (Thompson et al., 1990; Wong & Beasley, 1990). These weight restrictions may be used to reflect the relative importance of each input and output factor.

Second, the interpretation of the results can be judgmental. DEA does not give a single number that tells us that there is a true bias or not. One way that we can overcome this limitation is to compare the results to some benchmark scores depending on different characteristics like type of courses, the work load of a course, the grade distribution of course, gender, and level of the course (Morgan et al. 2003, Whitworth et al. 2002). Therefore, the collection of additional data or the impacts of exogenous factors will add additional insights.
Another limitation of the approach developed here is the possible biases and validity associated with the ‘halo effect’. The technique does not capture the possibility of students responding positively based on ‘output’ perceptions and adjusting their inputs accordingly. But, some of this has been mitigated by having mixed question types (positive and negative) in the survey questionnaire. For example, instead of a question stating “This instructor was effective” (on a 1 to 4 scale) it would be stated in the negative “This instructor was ineffective” to help prevent some of this systemic bias.

By applying DEA, administrators and faculty can decrease the discrepancy between them with respect to using course evaluation as a means of measuring teaching effectiveness. For example, Morgan et al. (2003) found that accounting administrators believe student evaluations measure teaching effectiveness to a greater degree than faculty, while faculty members believe their personality is the primary determinant of ratings on student evaluations.

Even though DEA cannot solve all the problems related to the course evaluation, the method can definitely be used as supplements to the existing system to interpret the possible problems regarding the validity of evaluation instruments and/or make the interpretation of the course evaluation instrument more meaningful. As suggested by others, however, the best practice should be that student evaluation shouldn’t be used as the only measure for faculty teaching evaluation but use multiple criteria rather than just focus on one mechanism.

REFERENCES

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

Charnes A, Cooper WW, Rhodes E. 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research* 2: 429-444.

Crumbley, D. L., Henry, B. K., and Kratchman, S. H. 2001. Students' perceptions of the evaluation of college teaching, *Quality Assurance in Education*, 9 (4):197-207.

Crumbley, D. L., and Fliedner, E. 2002. Accounting administrators' perceptions of student evaluation of teaching (SET) information, *Quality Assurance in Education*, 10 (4): 213-222.

Hunt, S.C., Eaton, T.V., and Reinstein, A. 2009. Accounting Faculty Job Search in a Seller's Market, *Issues in Accounting Education* 24 (2), 157-185.

Morgan, D. A., Sneed, J. and Swinney, L. 2003. Are student evaluations a valid measure of teaching effectiveness: perceptions of accounting faculty members and administrators, *Management Research News*, 26 (7): 17-32.

Nimmer, J.G. and Stone, E.F. 1991. Effects of grading practices and time of rating on student ratings of faculty performance and student learning, *Research in Higher Education*, 32 (April): 195-215.

O'Toole, D. M., Spinelli, M. A., and Wetzel, J. N. 2000. The Important Learning Dimensions in the School of Business: A Survey of Students and Faculty, *Journal of Education for Business*, 75 (6): 338-340.

Petersen, R.L., Berenson, M.L., Misra, R.B., and Radosevich, V.J. 2008. An Evaluation of Factors Regarding Students' Assessment of Faculty in a Business School, *Decision Sciences The Journal of Innovative Education*, 6(2): 375-402.

Read, W. J., Rama, D. V., and Raghunandan, K. 2001. The Relationship Between Student Evaluations of Teaching and Faculty Evaluations. *Journal of Education for Business*, 76(4): 189-198.

Seiler, V.L. and Seiler, M. J. 2002. Professor who make the grade, *Review of Business*, 23 (2): 39-44.

Tang, T.L.P. 1997. Teaching Evaluation at a Public Institution of Higher Education: Factors Related to the Overall Teaching Effectiveness, *Public Personnel Management*, 25: 379-384.

Thompson, R.G., Langemeier, L.N., Lee, C.T., and Thrall, R.M. 1990. The Role of Multiplier Bounds in Efficiency Analysis with Application to Kansas Farming, *Journal of Econometrics,* 46 (1/2): 93-108.

Wachtel, H.K. 1998. Student evaluation of college teaching effectiveness: A brief review. *Assessment and Evaluation in Higher Education*, 23 (2): 191-211.

Whitworth, J. E., Price, B. A., and Randall, C. H. 2002. Factors that affect college of business student opinion of teaching and learning, *Journal of Education for Business*, 77 (5): 282-90.

Wong, Y-H B., and Beasley, J E. 1990. Restricting Weight Flexibility in Data Envelopment Analysis, *Journal of the Operational Research Society*, 41 (9): 829-835.

Appendix

Mathematical Formulation for Basic DEA Models

DEA productivity models for a given decision-making unit (DMU) can use ratios based on the amount of outputs (ratings) per given set of inputs (ratings). The definition of a DMU can vary greatly, from individuals (students) to classes to schools, as long as the unit can be modeled with input and output values. DEA allows for the simultaneous analysis of multiple inputs to multiple outputs, a multi-factor productivity approach. The general efficiency measure used by DEA is best summarized by equation (A1).

$$E_{ks} = \frac{\sum_y O_{sy} v_{ky}}{\sum_x I_{sx} u_{kx}} \tag{A1}$$

where:
(E_{ks}) is the efficiency or productivity measure of DMU s, using the weights of test DMU k;
(O_{sy}) is the value of output y for DMU s;
(I_{sx}) is the value for input x of DMU s;
(v_{ky}) is the weight assigned to DMU k for output y; and
(u_{kx}) is the weight assigned to DMU k for input x.

In the basic DEA ratio model developed by Charnes, Cooper, and Rhodes (1978) (CCR), the objective is to maximize the efficiency value of a test DMU k from among a reference set of DMUs s, by selecting the optimal weights associated with the input and output measures. The maximum efficiencies are constrained to 1. The formulation is represented in expression (A2).

maximize
$$E_{kk} = \frac{\sum_y O_{ky} v_{ky}}{\sum_x I_{kx} u_{kx}}$$

subject to:
$$\begin{aligned} & E_{ks} \le 1 \qquad \forall \quad \text{DMUs} \;\; s \\ & u_{kx}, v_{ky} \ge 0 \end{aligned} \tag{A2}$$

This nonlinear programming formulation (A2) is equivalent to formulation (A3) (see Charnes et al. (1978) for a complete transformation explanation):

maximize
$$E_{kk} = \sum_y O_{ky} v_{ky}$$

subject to:
$$\begin{aligned} & E_{ks} \le 1 \qquad \forall \quad \text{DMUs} \;\; s \\ & \sum_x I_{kx} u_{kx} = 1 \\ & u_{kx}, v_{ky} \ge 0 \end{aligned} \tag{A3}$$

The transformation is completed by constraining the efficiency ratio denominator from (A2) to a value of 1, represented by the constraint $\sum_x I_{kx} u_{kx} = 1$.

The result of formulation (A3) (the CCR formulation) is an optimal simple or technical efficiency value (E_{kk}*) that is at most equal to 1 (this formulation has also been defined as the constant returns to scale formulation). If $E_{kk}* = 1$, then no other DMU is more efficient than DMU k for its selected weights. That is, $E_{kk}* = 1$ has DMU k on the optimal frontier and is not dominated by any other DMU. If $E_{kk}* < 1$, then DMU k does not lie on the optimal frontier and there is at least one other DMU that is more efficient for the optimal set of weights determined by (A3). The formulation (A3) is executed s times, once for each DMU.

The dual of the CCR formulation (also defined as the envelopment side) is represented by model (A4):

minimize θ

subject to:

$$\begin{aligned} &\sum_s \lambda_s I_{sx} - \theta I_{sx} \le 0 && \forall \text{ Inputs } I \\ &\sum_s \lambda_s O_{sy} - O_{ky} \ge 0 && \forall \text{ Outputs } O \qquad \text{(A4)} \\ &\lambda_s \ge 0 && \forall \text{ DMUs } s \end{aligned}$$

The CCR model has an assumption of constant returns to scale for the inputs and outputs. To take into consideration variable returns to scale, a model introduced by Banker, Charnes, and Cooper (1984) (BCC) is utilized. The BCC model aids in determining the scale efficiency of a set of units (which is a technically efficient unit for the variable returns to scale model). This new model has an additional convexity constraint defined by limiting the summation of the multiplier weights (λ) equal to one, or:

$$\sum_s \lambda_s = 1 \qquad \text{(A5)}$$

The use of the CCR and BCC models together helps determine the overall technical and scale efficiencies of the DMU respondents and whether the data exhibits varying returns to scale.

Figure 1: Simple Graphical Example of DEA for evaluation purposes.

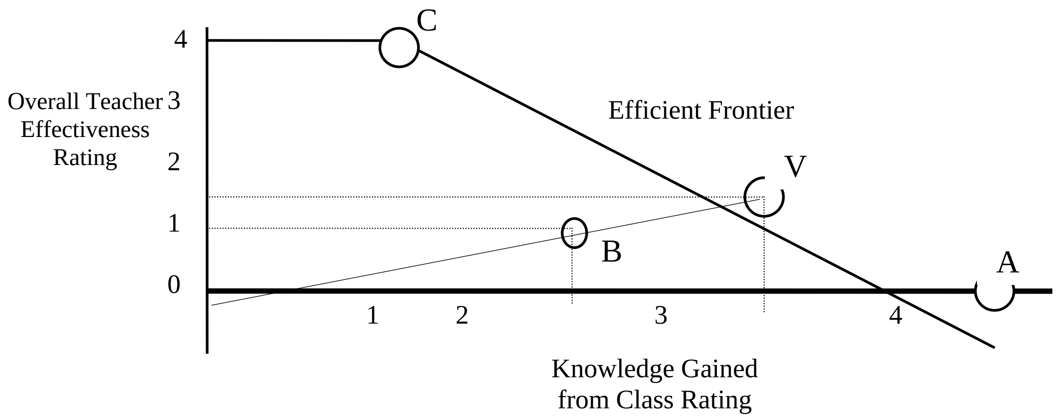

Figure 2: Graph of Constant Returns to Scale DEA Model for Student Evaluation Efficiency Scores.

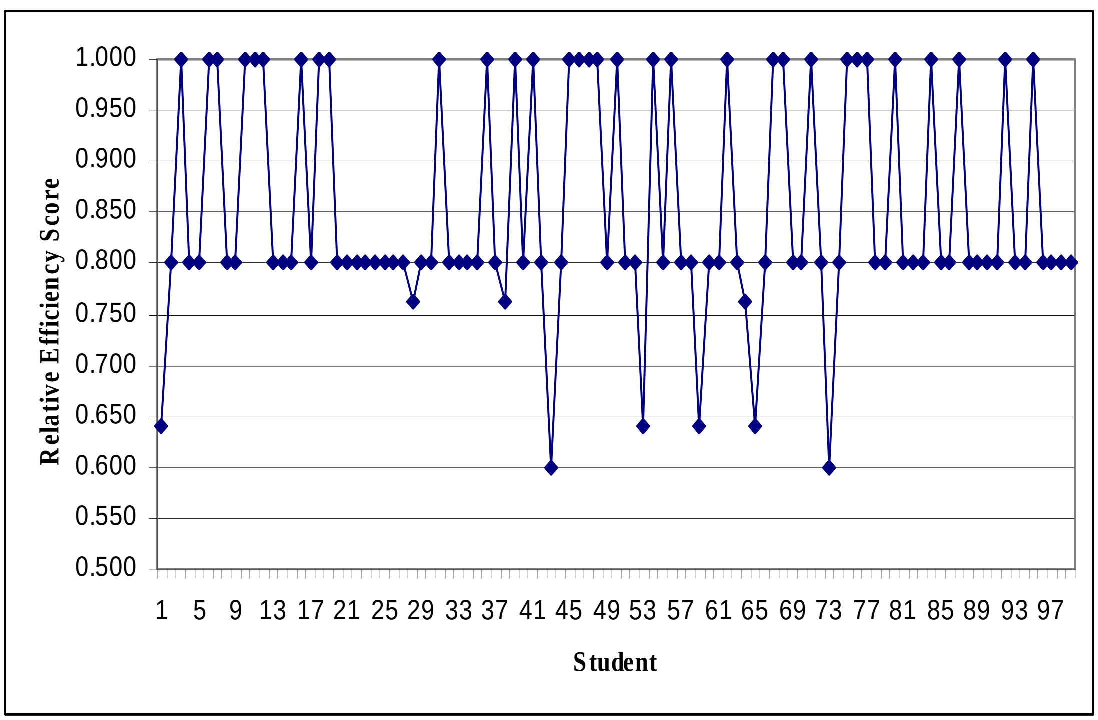

Table 1: The actual input factor scores as identified by the 99 students and model results.

Student	Input 1	Input 2	Input 3	Output 1	Output2	CCR*	BCC**
1	5	5	5	4	4	0.640	0.640
2	5	5	5	5	5	0.800	0.800
3	3	5	5	4	4	1.000	1.000
4	5	5	5	5	5	0.800	0.800
5	5	5	5	5	5	0.800	0.800
6	4	4	5	5	5	1.000	1.000
7	5	5	4	5	5	0.952	1.000
8	5	5	5	5	5	0.800	0.800
9	5	5	5	5	5	0.800	0.800
10	4	5	5	5	5	0.952	1.000
11	3	5	4	4	4	1.000	1.000
12	4	4	3	4	4	1.000	1.000
13	5	5	5	4	5	0.800	0.800
14	5	5	5	5	5	0.800	0.800
15	5	5	5	4	5	0.800	0.800
16	4	5	5	5	5	0.952	1.000
17	5	4	4	4	4	0.800	0.800
18	4	5	4	5	5	1.000	1.000
19	5	5	4	4	5	0.952	1.000
20	5	5	5	5	5	0.800	0.800
21	2	5	2	2	2	0.800	0.800
22	5	5	5	5	5	0.800	0.800
23	4	4	5	4	4	0.800	0.800
24	4	5	4	4	4	0.800	0.800
25	3	3	3	3	3	0.800	0.800
26	4	4	4	4	4	0.800	0.800
27	4	4	4	4	4	0.800	0.800
28	4	5	5	4	4	0.762	0.762
29	5	5	5	5	5	0.800	0.800
30	5	5	5	5	5	0.800	0.800
31	4	5	4	5	5	1.000	1.000
32	4	4	4	4	4	0.800	0.800
33	5	5	5	5	5	0.800	0.800
34	5	5	5	5	5	0.800	0.800
35	5	5	5	5	5	0.800	0.800
36	4	3	4	5	4	1.000	1.000
37	3	3	3	3	3	0.800	0.800
38	4	5	5	4	4	0.762	0.762
39	4	5	5	5	4	0.952	1.000
40	5	5	5	4	5	0.800	0.800
41	4	4	4	5	5	1.000	1.000
42	5	5	5	5	5	0.800	0.800
43	4	4	4	3	3	0.600	0.600
44	5	5	5	5	5	0.800	0.800
45	4	5	5	5	5	0.952	1.000
46	4	5	5	5	5	0.952	1.000
47	3	4	4	4	3	1.000	1.000
48	4	4	4	5	4	1.000	1.000
49	5	5	5	5	5	0.800	0.800
50	4	4	4	5	4	1.000	1.000
51	1	1	1	1	1	0.800	0.800

52	3	3	3	3	3	0.800	0.800
53	5	5	5	4	4	0.640	0.640
54	4	5	4	5	4	1.000	1.000
55	5	5	5	4	5	0.800	0.800
56	5	5	4	4	5	0.952	1.000
57	5	5	5	4	5	0.800	0.800
58	5	5	5	5	5	0.800	0.800
59	5	5	5	4	4	0.640	0.640
60	5	5	5	4	5	0.800	0.800
61	4	4	5	4	4	0.800	0.800
62	4	4	4	5	4	1.000	1.000
63	5	5	5	5	5	0.800	0.800
64	4	5	5	4	4	0.762	0.762
65	5	5	5	4	4	0.640	0.640
66	5	5	5	5	5	0.800	0.800
67	3	4	4	4	3	1.000	1.000
68	4	5	4	5	4	1.000	1.000
69	4	5	4	4	4	0.800	0.800
70	5	5	5	4	5	0.800	0.800
71	4	5	4	5	5	1.000	1.000
72	5	5	5	5	5	0.800	0.800
73	4	4	4	3	3	0.600	0.600
74	5	5	5	5	5	0.800	0.800
75	4	4	5	5	4	0.972	1.000
76	4	5	5	5	5	0.952	1.000
77	5	5	3	3	4	1.000	1.000
78	5	5	5	5	5	0.800	0.800
79	5	5	5	5	5	0.800	0.800
80	3	4	4	4	4	1.000	1.000
81	5	5	5	5	5	0.800	0.800
82	5	5	5	5	5	0.800	0.800
83	5	5	5	5	5	0.800	0.800
84	4	5	4	5	5	1.000	1.000
85	5	5	5	5	5	0.800	0.800
86	5	5	5	5	5	0.800	0.800
87	4	5	5	4	5	0.952	1.000
88	4	5	4	4	4	0.800	0.800
89	5	5	5	5	5	0.800	0.800
90	5	5	5	4	5	0.800	0.800
91	5	5	5	5	5	0.800	0.800
92	4	5	5	5	5	0.952	1.000
93	5	5	5	5	5	0.800	0.800
94	5	5	5	5	5	0.800	0.800
95	4	4	4	5	4	1.000	1.000
96	5	5	5	5	5	0.800	0.800
97	5	5	5	5	5	0.800	0.800
98	5	5	5	5	5	0.800	0.800
99	5	5	5	5	5	0.800	0.800

* Charnes, Cooper, and Rhodes (1978) Model
** Banker, Charnes, and Cooper (1984) Model
See the appendix for model descriptions.

Table 2(a): Cluster Distribution

		N	% of Total
Cluster	1	6	6.1%
	2	61	61.6%
	3	32	32.3%
Total		99	100.0%

Table 2(b): Clusters and centroids for DEA relative efficiency scores for student evaluation data

		CCR*			BCC**		
		Mean	Std. Deviation	Coef. of Variation***	Mean	Std. Deviation	Coef. of Variation***
Cluster	1	.62667	.020656		.62667	.020656	
	2	.79813	.008306		.79813	.008306	
	3	.98276	.022865		1.0000	.000000	
	Combined	.84742	.103473	.1221035	.85299	.110139	.129121

* Charnes, Cooper, and Rhodes (1978) Model
** Banker, Charnes, and Cooper (1984) Model
***The coefficient of variation statistic can be determined by dividing a sample's standard deviation by its mean.