Öneri C.6 S.22
Haziran 2004 ss.203-213

DEVELOPING EVALUATION CRITERIA OF ACADEMICIANS' TEACHING PERFORMANCE: DISCRIMINANT AND LOGISTIC REGRESSION APPLICATIONS

E. Serra YURTKORU[1], Beril SİPAHİ[2]

[1]*Marmara Üniversitesi, İ.İ.B.F., İngilizce İşletme Bölümü, Öğretim Görevlisi Dr.*
[2]*Marmara Üniversitesi, İ.İ.B.F., İngilizce İşletme Bölümü, Öğretim Görevlisi Dr.*

DEVELOPING EVALUATION CRITERIA OF ACADEMICIANS' TEACHING PERFORMANCE: DISCRIMINANT AND LOGISTIC REGRESSION APPLICATIONS

***Abstract:** Since 1997 students' course evaluation questionnaire (SCEQ) have been conducted in Business Administration Department of Marmara University to evalute teaching effectiveness of instructors. In this study data obtained from 2000-2001 and 2001-2002 questionnaires are used to conduct discriminant and logistic regression analysis. As a result of these analyses evaluation criteria of academicians' teaching performance are developed Findings of both the disciriminant and logistic regression methods show similar results. In both methods "satisfaction with the way course is conducted" is the most powerful item that discriminates effective instructors from ineffective instructors. This item is followed by "instructors being competent in his/her field" and "exams being a good measuring devise".*

***Keywords:** Discriminant Analysis, Logistic Regression Analysis, Course Evaluation, Student Evaluation, Instructor Performance*

DEVELOPING EVALUATION CRITERIA OF ACADEMICIANS' TEACHING PERFORMANCE: DISCRIMINANT AND LOGISTIC REGRESSION APPLICATIONS

***Özet:** Marmara Üniversitesi İİBF İngilizce İşletme bölümü 1997'den bu yana öğretim üyelerinin performanslarını değerlendirmek amacıyla öğren-cilerine ders değerlendirme anketi uygulanmaktadır. Bu çalışmada 2000-2001 ve 2001-2002 Eğitim ve Öğretim yılında yapılmış anketlerden elde edilen verilere diskriminant ve lojistik regresyon analizleri uygulanmış ve bu analizlerin sonucunda öğrencilerin bakış açıları esas alınarak öğretim üyesi performans değerlendirme kriterleri belirlenmiştir. Araştırma sonucunda hem diskriminant hem de lojistik regresyon metodunda benzer bulgular elde edilmiştir. Her iki analiz bulguları da "öğrencinin dersin işlenişinden memnun olması"nı yeterli ve yetersiz performansı ayırt edici en önemli unsur olarak göstermektedir. Bunu, "öğretim üyesinin konusuna hakim olması" ve "dersin sınavının iyi bir ölçüt" olması takip etmektedir.*

***Anahtar Kelimeler:** Diskriminant Analizi, Lojistik Regresyon Analizi, Ders Değerlendirme, Öğretim Üyesi Performansı*

I. INTRODUCTION

Performance Evaluation, which in the light of having assessed any member's past, current task performance, generates and implements actions that attempt to either reinforce or correct that level of performance [1]. Most important use of performance evaluation to those who are being evaluated is that it provides them feedback about what they are doing right and wrong and draws their attention to strengths and weaknesses exhibited in previous performance [2; 3].

Just like in any organization, performance evaluation is an important tool in higher education. The majority of universities use student evaluations as the most important measure of teaching effectiveness as part of the faculty member's performance evaluation for the past 70 years [4; 5; 6].

A key issue in performance evaluation is determining what constitutes valid criteria or measures of effective performance [2]. In our study we want to identify those criteria that are perceived by the students as valid measures of effective performance by their instructors.

In the following section, we will first review the student evaluation of teaching literature. Then we will introduce our study where we have conducted discriminant and logistic regression analyses to develop evaluation criteria for instructor teaching performance. After discussing the implications of our findings, we will present some ideas for further research.

II. LITERATURE REVIEW

Student evaluations of instructor performance are commonly used with the prime purpose of improving course and teaching quality. This tool is often used as part of the evaluation process for staff appraisal as well [7; 8].

If these instruments are used without alternative or collaborative measures, then students become the primary determinant of lecturers' success or failure in their academic career [5]. Few faculty members question the usefulness of ratings in providing feedback about teaching that can result in improved instruction, but many continue to challenge student rating in use of making personnel decisions regarding promotion, tenure, and merit process [9; 7;10].

Concerns regarding student evaluations that are discussed in the literature cover a) students lack the maturity and expertise to make judgement about course content or instructor style; b) students' ratings are measures of popularity rather than of ability; c) the rating forms themselves are both unreliable and invalid; d) other variables (such as grades received from the instructor, class size, or whether the course was compulsory or elected; and instructor characteristics) affect student ratings [9].

Much of the dissatisfaction surrounding student evaluations is based on the concern that student evaluations are more personality contests than they are valid measures of teaching effectiveness [6]. But it should be kept in mind that students are the only source of information about the learning environment, including teachers' ability to motivate students for continued learning, and rapport. Also student ratings encourage communication between students and their instructor which can raise the level of teaching [9;10]. In fact students are the most logical evaluators of the quality, the effectiveness of, and satisfaction with course content, method of instruction, textbook, homework, and student interest. [9].

Despite the fact that researchers have debated the reliability and validity of student evaluations of teaching, all authors recommended that student evaluation should be used [11]. Actually reliability and validity concerns are particularly related to the summative purpose of student evaluation ratings and its use as an input to personnel decisions [12].

However discussions on summative purpose of student evaluation, using these evaluations with the intent of determining instructors' tenure and promotion, is not relevant for our study. Student evaluation questionnaire we use is designed with formative purpose to help our department members to improve and enhance their teaching skills.

In fact several past studies found that student evaluation of teaching offers a reliable and valid assessment of instructors [13; 14; 12; 15]. Research results have shown that the student evaluations are reliable and as long as the instructor does not change the teaching style, the evaluation ratings are consistent [16]. One more important finding is that the number of students evaluating the instructor is positively correlated with reliability [16].

As noted earlier key issue in performance evaluation is determining valid criteria of performance. The present study attempts to enlighten the criteria or measures of effective instructor performance as perceived by the students. Therefore we will try to discriminate instructors' performance from the students point of view.

III. METHOD

The students' course evaluation questionnaire (SCEQ) is given to undergraduate students for all courses at the end of each semester since 1997 at Marmara University Faculty of Economics and Administrative Sciences - Business Administration Department.

III.1 Sample

In this study we used the data of SCEQ from 2000-2001 and 2001-2002 semesters.

The sample consists of 5996 questionnaires in total. In this study 70 % of the data set is used for constructing a model (analysis sample) and 30 % of the data set is used for testing the model (holdout sample). For selecting analysis sample we used random sampling technique with Bernoulli distribution. As a result our analysis sample consists of 3033 observations and holdout sample consists of 2963 observations.

III.2 Instrument

SCEQ* has 22 items where students' are asked to rate how much they agree with each item on a 5 point scale where "not at all" equals 1 and " definitely " equals 5. Items can be found in Table 1.

There is also one nominal scale question where students' answer their willingness to take another course from the same instructor. This question is used to distinguish the effective instructor from ineffective ones from the students' perspective.

III.3 Analysis

Discriminant analysis and logistic regression are multivariate statistical techniques that can be used to

*The instrument SCEQ used to measure students' ratings of courses is developed by Prof. Dr. Suna TEVRUZ and Dr. Serra YURTKORU.

predict nonmetric dependent variable with a set of metric independent variables. [17;18]

There are many similarities between the two methods. Discriminant analysis and logistic regression are used with the purpose of profiling, differentiation, and classification [17, 18].

Table 1: Students' course evaluation questionnaire (SCEQ)

	SCEQ ITEMS
Q1	I attend the course regularly (Students' self-evaluation)
Q2	Instructor attends regularly
Q3	Instructor comes to class on time
Q4	Instructor makes use of whole class period
Q5	Instructor lectures using real life examples
Q6	Instructor does not lecture in a logical sequence
Q7	Instructor gives the lecture by reading from notes/books
Q8	Instructor gives pop quizzes
Q9	Students are asked to prepare homework/ project/ research
Q10	Instructor encourages students' participation
Q11	Instructor speaks English fluently
Q12	Instructor is competent in his/her field
Q13	I am satisfied with the way course is conducted
Q14	I found this course very useful
Q15	Text book of this course is very useful in learning the content
Q16	The language of the text book is understandable
Q17	Instructor follows the syllabus
Q18	Exam of this course encourages creativity
Q19	Exam of this course requires application of content
Q20	Exam of this course requires memorisation of content
Q21	Exam of this course is a good measuring devise
Q22	Exam time of this course is suffice

Discriminant analysis has the underlying assumptions of multivariate normality, whereas logistic regression analysis has no such distributional assumptions [19]. Equal covariance matrices in the two groups; is also required for the prediction rule to be optimal in discriminant analysis, which is not required in logistic regression [20].

A difference between the techniques is that logistic regression is also available in situations where independent variables are nonmetric [17; 19], since in our study this is not the case we will not get into details of this property.

When the basic assumptions are met both techniques give comparable predictive and classificatory results and employ similar diagnostic measures [17].

Logistic regression is an attractive alternative to discriminant analysis when the assumptions are violated; discriminant analysis should be used when the multivariate normal assumptions are met, because discriminant analyis is computationally more efficient [17; 19]. Therefore the choice between the two techniques depends on whether the assumptions are met or not [19].

In this study our purpose was to find out how groups differ in terms of the values of the independent variables: to profile dependent variable. Discriminant and logistic regression analyses were conducted with SPSS package program to profile effective and ineffective instructors.

IV. FINDINGS

First discriminant analysis is explained then logistic regression, and then the results of the both techniques are compared.

IV.1 Discriminant Analysis

Before starting a discriminant analysis principal assumptions should be diagnosed. The basic assumptions underlying discriminant analysis can be discussed under two categories: assumptions related to formation of discriminant function which are normality, linearity, and multicollinearity, and assumption related to estimation of discriminant function which is equal covariance matrices [21].

In order to diagnose normality assumption we conducted Kolmogorov-Smirnov with Lilliefors significance correction and Normal Q-Q Plots to our independent variables and found out that normality assumptions are met. To test the multicollinearity we analyzed correlations between the independent variables and reached the conclusion that multicollinearity does not exist since all correlation coeficients were low in magnitude[22]. As a result in our study assumptions related to formation of discriminant function are met at acceptable levels. Normality and multicollinearity results are not presented in tables because of the limited space provided.

Table 2: Tests of equality of covariance matrices

Box M	517.312
F value	14.304
df1	36
df2	8376516
p value	0.000

The most common test for equality of covariance matrices is the Box's M test [17]. The result of the test can be seen in Table 2. As can be seen from the table the assumption of equal covariance matrices were violated.

Violation in this assumption affects the estimation of discriminant function but not the formation. Also Klecka emphasises that the most difficult assumption to meet is equal covariance matrices, but discriminant analysis is a rather robust technique which can tolerate this assumption [19]. Also in our analysis hit ratio of the model is very satisfactory (80%) which indicates this violation did not effect the estimation power.

Table 3: Group Means and Standard Deviations of Independent Variables

	Effective Instructor Performance n=1289		Ineffective Instructor Performance n=746		Total N=2035	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Q1	4.136	1.014	3.621	1.230	3.947	1.126
Q2	4.432	0.801	3.975	1.055	4.264	0.929
Q3	4.335	0.845	3.843	1.101	4.155	0.976
Q4	4.263	0.859	3.692	1.142	4.054	1.010
Q5	4.267	0.833	3.319	1.209	3.919	1.088
Q6	2.481	1.212	2.968	1.132	2.660	1.206
Q7	2.892	1.353	3.168	1.343	2.993	1.356
Q8	2.056	1.416	1.743	1.194	1.941	1.347
Q9	2.938	1.553	2.511	1.455	2.781	1.531
Q10	4.125	0.966	3.139	1.295	3.764	1.196
Q11	4.195	0.861	3.239	1.239	3.844	1.115
Q12	4.383	0.709	3.370	1.146	4.012	1.018
Q13	3.950	0.967	2.432	1.274	3.393	1.312
Q14	3.968	0.958	2.778	1.223	3.532	1.208
Q15	3.859	1.071	3.158	1.290	3.602	1.204
Q16	3.971	0.998	3.343	1.268	3.741	1.145
Q17	4.192	0.891	3.409	1.222	3.905	1.092
Q18	3.405	1.234	2.355	1.232	3.020	1.333
Q19	3.933	1.057	3.001	1.347	3.592	1.254
Q20	3.514	1.140	3.645	1.312	3.562	1.207
Q21	3.632	1.065	2.458	1.165	3.202	1.239
Q22	3.817	1.108	3.067	1.273	3.542	1.225

Since Box's M was significant, we run a second analysis to see whether using a separate-groups covariance matrix changed the classification. The classification results have not changed, so it was not worth using separate covariance matrices. Therefore we continued with within group covariance matrix. Box's M can be overly sensitive to large data files [23] which was likely what happened in our study. Hence we continued to perform discriminant analysis.

Group means and standard deviations of independent variables are given in Table 3. We can identify the variables with the largest differences in the group means by looking at the below table. Since discriminant analysis excludes missing values the sample size has decreased to total of 2035.

Table 4 shows the Wilks' Lambda and ANOVA results of test for the equality of group means. These tests indicate that all of the independent variables have significant univariate differences between the two groups.

When the objective of a study is to determine the discriminating capabilities of the entire set of benefits, all variables would be entered into the model simultaneously. Since we aimed to determine which variables are the most efficient in discriminating between effective and ineffective instructor performances, a stepwise procedure is used.

Table 4: Test for the Equality of Group Means

Independent Variables	Wilks' Lambda	F value	d.f.1	d.f.2	p value
Q1	0.950	103.920	1	2033	0.000
Q2	0.940	121.550	1	2033	0.000
Q3	0.940	127.630	1	2033	0.000
Q4	0.930	163.230	1	2033	0.000
Q5	0.820	435.380	1	2033	0.000
Q6	0.960	79.970	1	2033	0.000
Q7	0.990	19.680	1	2033	0.000
Q8	0.990	25.860	1	2033	0.000
Q9	0.980	37.450	1	2033	0.000
Q10	0.840	380.440	1	2033	0.000
Q11	0.830	418.650	1	2033	0.000
Q12	0.770	607.120	1	2033	0.000
Q13	0.690	917.410	1	2033	0.000
Q14	0.770	593.010	1	2033	0.000
Q15	0.920	173.650	1	2033	0.000
Q16	0.930	152.830	1	2033	0.000
Q17	0.880	275.320	1	2033	0.000
Q18	0.860	342.420	1	2033	0.000
Q19	0.870	299.130	1	2033	0.000
Q20	1.000	5.590	1	2033	0.018
Q21	0.790	535.260	1	2033	0.000
Q22	0.910	193.800	1	2033	0.000

Table 5: Result of Stepwise Discriminant Analysis

Step	Variable Entered	Min. D^2	Wilks' Lambda	F value	p value	df1	df2	df3
1	Q13	1.941	0.689	917.41	0.000	1	1	2033
2	Q21	2.175	0.664	513.62	0.000	2	1	2033
3	Q12	2.348	0.647	369.40	0.000	3	1	2033
4	Q6	2.401	0.642	283.20	0.000	4	1	2033
5	Q14	2.441	0.638	230.28	0.000	5	1	2033
6	Q1	2.468	0.635	193.89	0.000	6	1	2033
7	Q5	2.486	0.634	167.32	0.000	7	1	2033
8	Q22	2.500	0.633	147.13	0.000	8	1	2033

Table 5 provides the overall stepwise discriminant analysis results after all the significant variables have been included in the estimation of the discriminant function. This summary table describes the eight variables that were significant discriminators based on their Wilks' Lambda and minimum Mahalonobis D^2 values (Q13: I am satisfied with the way course is conducted, Q21: Exam of this course is a good measuring devise, Q12: Instructor is competent in his/her field, Q6: Instructor does not lecture in a logical sequence, Q14: I found this course very useful, Q1: I attend the course regularly, Q5: Instructor lectures using real life examples, Q22: Exam time of this course is suffice, respectively).

Table 6: Canonical Discriminant Function

Function	Eigen Value	% of variance	Cumulative %	Canonical Correlation	Wilks' lambda	Chi-square	df	p value
1	0.581	100	100	0.606	0.633	929.372	8	0.000

Table 7: Canonical Discriminant Function Coefficients

Independent Variables	Standardized Coefficients	Unstandardized Coefficients
Q1	0.098	0.090
Q5	0.097	0.098
Q6	-0.141	-0.119
Q12	0.229	0.256
Q13	0.480	0.440
Q14	0.132	0.124
Q21	0.242	0.219
Q22	0.080	0.068
Constant		-4.323

Our canonical discriminant function was significant at $\alpha=0.001$ and canonical correlation was 0.606 (See Table 6). Standardized and unstandardized coefficients of the discriminant function are displayed in Table 7. We will use the unstandardized coefficients to determine the discriminant scores that will be used in classification. While the unstandardized coefficients do tell us the absolute contribution of a variable in determining the discriminant score, this information may be misleading when the meaning of one unit change in the value of a variable is not the same from one variable to another [21] .

When we want to know the relative importance of the variable, we need to look at the standardized coefficients. In our findings Q13, *I am satisfied with the way course is conducted*, was the variable with the highest contribution to the discriminant function. This was followed by Q21 and Q12 (Instructor is competent in his/her field; Exam of this course is a good measuring devise respectively). There was only one variable with negative contribution Q6 *Instructor does not lecture in a logical sequence.*

Table 8: Classification Function Coefficients (Fisher's linear discriminant functions)

Independent Variables	Group 1 Effective Instructor Performance	Group 2 Ineffective Instructor Performance
Q1	2.222	2.080
Q5	2.010	1.855
Q6	2.404	2.593
Q12	3.414	3.010
Q13	0.079	-0.617
Q14	0.825	0.629
Q21	0.229	-0.117
Q22	1.389	1.281
Constant	-24.901	-18.399

The classification function coefficients also known as Fisher's linear discriminant functions are given in Table 8 and used for classification purposes. Groups' centroids are also reported and they represent the mean of the individual discriminant function scores for each group (See Table 9).

Table 9: Group Means (Centroids) of Canonical Discriminant Function

Group	Group Means
Effective Instructor Performance	0.580
Ineffective Instructor Performance	-1.001

Unstandardized canonical discriminant functions evaluated

Group centroids can be used to interpret the discriminant function results from an overall perspective. Table 9 reveals that the group centroid for the effective instructor performance (Group 1) is 0.580, whereas group centroid for the ineffective instructor performance (Group2) is –1.001. Critical cutting score was calculated from the group centroids by taking their average (Zc=0.2895). The graphical illustration of the plot of centroids can be found in Figure 1.

To determine how well our discriminant function can predict group memberships, a classification matrix was formed. The percentage of the known cases, which are correctly classified, is an additional measure of group differences. We can use it along with the overall Wilks' Lambda and the canonical correlations to indicate the amount of discrimination contained in the variables [21]. The proportion of cases correctly classified indicates the accuracy of the procedure and indirectly confirms the degree of group separation. In Table 10 we portrayed the classification matrix result of analysis sample.

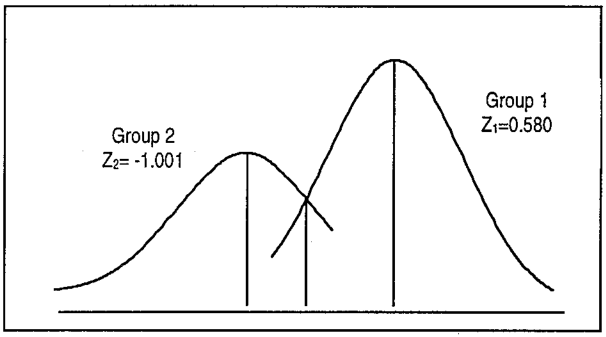

Figure1: Plot of Group Centroids (Z)

The hit ratio of our analysis sample was 80.2 %. Hit ratio was calculated by dividing the number correctly classified cases to the total number of cases [(1964 + 1133)/3862=0.802]

Table 10: Classification Result: Analysis Sample

Actual Group Membership	Analysis Sample Predicted Group Membership Effective Instructor Performance	Ineffective Instructor Performance	Total
Effective Instructor Performance	1964	404	2368
	82.9%	17.1%	100 %
Ineffective Instructor Performance	361	1133	1494
	24.2%	75.8%	100%
Cases Correctly Classified 80.2 %			

This hit ratio is obtained from analysis sample may tend to overestimate the power of classification procedure because the validation is based on the same cases used to derive classification functions .

Therefore we have applied our discriminant function to the holdout sample. The hit ratio of the classification of holdout sample was 81.5 % (See Table 11).

The predictive accuracy of the discriminant function is measured by the hit ratio however to determine an acceptable level of predictive accuracy we calculated chance base criteria. In our study since we have two groups the chance criteria is 0.50 %. Hence our classification accuracy of 80.2 % and 81.5% were substantially higher than the chance criterion of 50 %.

Table 11: Classification Results: Holdout Sample

Actual Group Membership	Holdout Sample Predicted Group Membership Effective Instructor Performance	Ineffective Instructor Performance	Total
Effective Instructor Performance	854	160	1014
	84.2%	15.8%	100%
Ineffective Instructor Performance	149	508	657
	22.7%	77.3%	100%
Cases Correctly Classified 81.5 %			

A statistical test for the discriminatory power of the classification matrix when compared with a chance model is Press's Q statistic [17]. This simple measure compares the number of correct classifications with the total sample size and the number of groups.

The Press's Q statistic is calculated by the

$$\text{Press's Q} = \frac{[N - (nK)]^2}{N(K-1)}$$

following formula:

where
N= total sample size
n= number of observations correctly classified
K= number of groups

The Press's Q values of analysis sample and holdout sample were 1408.14 and 663.56 respectively. In both instances, the calculated values exceeded the critical $\chi 2$value of 6.63 (α=0.01) indicating we have significantly high hit ratios.

In order to estimate our model we then performed logistic regression.

IV.2 Logistic Regression

Logistic regression presents an alternative to discriminant analysis that may be preferred by many researchers due to its similarity to multiple regression analysis [17].

Logistic regression is estimated much like multiple regression in that a base model is first estimated to provide a standard for comparison. Only this time log likelihood value is calculated instead of sum of squares to selecting parameters [24]. Table 12 contains the base model result for the logistic regression analysis where, the log likelihood value (-2LL) is 2676.275.

Table 12: Logistic Regression Base Model Results

Iteration	-2 Log likelihood	Coefficients Constant
Step 1	2676.275	0.548

a. Constant is included in the model.
b. Initial -2 Log Likelihood: 2676.275

Since we aimed to identify which variables are the most efficient predictors between effective and ineffective instructor performances, as in the discriminant analysis, a stepwise procedure is used.

The score statistics, a measure of association used in logistic regression, along with the partial correlation for each independent variable are indicators of the variable selected in the stepwise procedure. Several criteria can be used to guide entry: greatest reduction in the –2LL value, greatest Wald coefficient, or highest conditional probability. [17]. Here we employed the likelihood ratio test since this tests reestimates the model with each variable eliminated and judges the change in the log likelihood every time [20].

Table 13 provides the stepwise logistic regression analysis results after all the significant variables have been included in the estimation of the logistic regression equation. This summary table describes the eight variables that were significant based on Wald statistics values (Q1: I attend the course regularly, Q5: Instructor lectures using real life examples, Q6: Instructor does not lecture in a logical sequence, Q12: Instructor is competent in his/her field, Q13: I am satisfied with the way course is conducted, Q14: I found this course very useful, Q21: Exam of this course is a good measuring devise, and Q22: Exam time of this course is suffice, respectively).

Table 13: Variables in the equation

Variables	B	S.E.	Wald	df	p value	R	Exp(B)
Q1	0.123	0.055	4.954	1	0.026	0.033	1.131
Q5	0.168	0.069	5.920	1	0.015	0.038	1.182
Q6	-0.205	0.056	13.377	1	0.000	-0.065	0.814
Q12	0.417	0.079	27.790	1	0.000	0.098	1.518
Q13	0.542	0.066	67.398	1	0.000	0.156	1.719
Q14	0.188	0.067	7.946	1	0.005	0.047	1.207
Q21	0.341	0.062	30.347	1	0.000	0.103	1.406
Q22	0.131	0.054	5.897	1	0.015	0.038	1.140
Constant	-5.574	0.432	66.388	1	0.000		

Discriminant analysis is a linear function that can be used discriminate between different values of the dependent variable. Logistic regression is a nonlinear function that describes the probability of choice between two alternatives of the independent variable [18]. Table 13 provides the estimated coefficients of the logistic regression model. Using the B coefficients the logistic regression equation for the probability of event to occur can be calculated as

$$\text{Probability (Y)} = \frac{1}{1 + e^{-Z}}$$

where $Z = -5.574 + 0.123(\mathbf{Q1}) + 0.168\,(\mathbf{Q5}) - 0.205\,(\mathbf{Q6}) + 0.417\,(\mathbf{Q12}) + 0.542\,(\mathbf{Q13}) + 0.188\,(\mathbf{Q14}) + 0.341\,(\mathbf{Q21}) + 0.131\,(\mathbf{Q22})$.

In general if the estimated probability of the event is less than 0.5, it is predicted that the event will not occur. If the probability is greater than 0.5 it is predicted that the event will occur. [20].

When we want to assess the contribution of each variable in logistic regression, we need to look at the partial correlation R and Exp (B) values (See Table 13).

When the partial correlation R's analyzed Q13, *I am satisfied with the way course is conducted*, was the variable with the highest contribution to the logistic regression model(R= 0.156). This was followed by Q21 and Q12 (*Instructor is competent in his/her field*, R=0.103; *Exam of this course is a good measuring devise*, R=0.098 respectively). There was only one variable with negative contribution Q6 *Instructor does not lecture in a logical sequence* (R=–0.065) meaning increase in the value of this variable would decrease the likelihood of the event to occur.

Exp (B) is a value by which the odds ratio of an event (ratio of the probability that an event will occur to probability that it will not) change when an independent variable increases by one unit [20; 24]. In our study, parallel to partial R results, Q13, *I am satisfied with the way course is conducted*, was the variable with the highest contribution to the logistic regression model (Exp (B)=1.719). If the value is greater than 1, the odds are increased. This was followed by Q12 and Q21 (*Instructor is competent in his/her field*, Exp (B)= 1.518; *Exam of this course is a good measuring devise*, Exp (B)= 1.406 respectively). There was only one variable with Exp (B) value less than 1 which would decrease the odds: Q6 *Instructor does not lecture in a logical sequence* (Exp (B)= 0.814).

In assessing model fit several measures are available. One of them is evaluating the –2LL where smaller–2LL value indicating better model fit. The –2LL value was reduced from the base model value of 2676.275 to 1808.002 in our study. Model and improvement chi-square values can be found in Table 14.

Table 14: Model and Improvement Chi-Square Summary

	Chi-square	d.f.	p value
Model	868.273	8	0000
Improvement	4946	1	0026

In addition to the statistical chi-square tests, several R^2 measures can be used to evaluate model fit in logistic regression, which resemble the R^2 in the multiple regression models. Overall model fit statistics are given in Table 15.

Table 15: Overall Model Fit

-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square	Hosmer and Lemeshow Test		
			Chi-square	d.f.	p value
1808.002	0.347	0.475	14.515	8	0.069

The Cox and Snell R^2 measure operates in the same manner as R^2 in the multiple regression, with higher values indicating greater model fit. However this measure is limited in that it can not reach maximum value of 1, so Nagelkerke's modification that have the range 0 to 1 can be used [17]. In our study the Cox and Snell value is 0.347 and the Nagelkerke value is 0.475.

The third measure the "pseudo R^2" based on the improvement in the –2LL value can also be calculated as:

$$R^2_{\log it} = \frac{-2LL_{null} - (-2LL_{model})}{-2LL_{null}}$$

The pseudo R^2 value of our model is found to be 0.324.

The final measure of the model fit is Hosmer and Lemeshow value, which measures the correspondence of the actual, and the predicted values of the dependent variable [25]. Better model fit is indicated by smaller difference between the observed and predicted classification. A good model fit is indicated by nonsignificant chi-square value [17]. In our study Homer and Lemeshow significance value is 0.069 indicating no difference in the distribution of the actual and predicted dependent values (See Table 15).

To determine how well our logistic model can predict group memberships, a classification matrix was formed identical in nature to those used in discriminant analysis. In Table 16 we illustrated the classification matrix result of analysis sample. The hit ratio of our analysis sample was 80.4 %.

Table 16: Classification Result: Analysis Sample

Analysis Sample	Predicted Group Membership		
Actual Group Membership	Effective Instructor Performance	Ineffective Instructor Performance	Total
Effective Instructor Performance	1147	144	1291
	88.8%	11.2%	100 %
Ineffective Instructor Performance	256	490	746
	34.3%	65.7%	100%
Cases Correctly Classified 80.4 %			

Table 17: Classification Results : Holdout Sample

Holdout Sample			
	Predicted Group Membership		
Actual Group Membership	Effective Instructor Performance	Ineffective Instructor Performance	Total
Effective Instructor Performance	530	48	578
	91.7%	8.3%	100%
Ineffective Instructor Performance	117	189	306
	38.2%	61.8%	100%
Cases Correctly Classified 81.3 %			

As we discussed in the previous section hit ratio obtained from analysis sample may tend to overestimate the power of classification procedure because the validation is based on the same cases used to derive classification functions. Therefore we applied our logistic regression model to the holdout sample. The hit ratio of the classification of holdout sample was 81.3 % (See Table 17).

The individual hit ratios are consistently high and do not indicate problem in predicting in either of the two groups.

Finally the histogram of estimated probabilities can be found in Figure 2. When the model distinguishes the two groups, the cases for which the event has occurred should be to the right side of 0.5, while those cases for which the event has not occurred should be to the left of 0.5. [20]. Here we see there are some misclassifications yet our model is more successful in predicting the effective instructor performance, which is supported by the classification results as well.

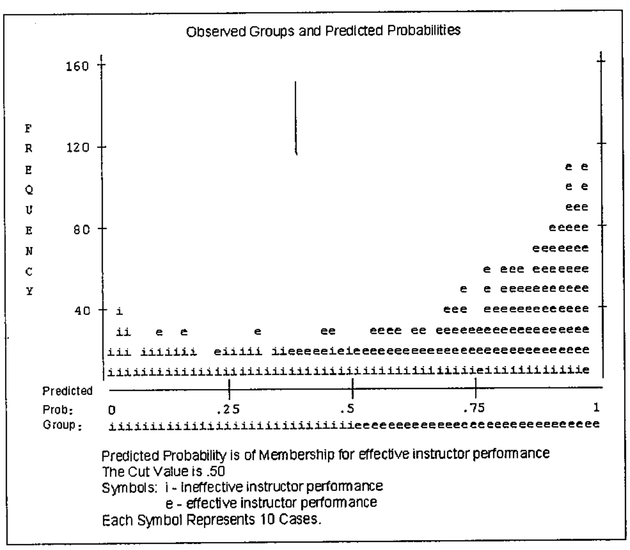

Figure 2: Histogram of Estimated Probabilities

IV.3 Comparison of Techniques and Findings

To determine valid criteria of effective instructor performance we conducted two separate multivariate techniques.

Discriminant analysis was performed to constitute a linear function that was used to discriminate between different values of the dependent variable: effective instructor performance and ineffective instructor performance. Whereas logistic regression was performed to constitute a nonlinear function that described the probability of choice between two alternatives of the dependent variable.

As we discussed before application of discriminant analysis requires some assumptions. In our study assumptions related to formation of discriminant function were met, but assumption related to estimation of discriminant function, which is equal covariance matrices was not satisfied. Violation in this assumption affects the estimation of discriminant function but not the formation.

Note that discriminant analysis can be used for exploration or for testing [18]. Since our aim was developing criteria to feedback to instructors we still performed the analysis.

The hit ratios of the model and holdout sample (80.2, 81.5 respectively) indicated that this violation has not effected the estimation power and supported our decision of continuing. Just as Klecka [19] emphasised our discriminant analysis was rather robust and could tolerate this assumption.

When we compare the two techniques' goodness of fit results and accuracy of classification, we see that both provided similar findings. Likewise the significant variables derived from both methods were same. Therefore for this study both methods' contributions were same and could be used interchangeably.

V. CONCLUSION AND DISCUSSION

Studies on students' evaluation of teaching mainly deal with the use of these evaluations as a decision making tool. Hence they are used for affecting instructors pay increases, retention and promotion. Probably this is the main reason why their validity is under investigation and results from different studies only support each other at broader levels.

However our study has a formative purpose to help our department members to improve and enhance their teaching skills. Even the student evaluation questionnaire we use, SCEQ, is designed with this purpose.

The present study aimed to enlighten the criteria of effective instructor performance. The key issue in performance evaluation is determining valid criteria of performance and students are the only source of instructor performance of teaching. Therefore we believe that by incorporating students' perception the criteria developed will be trusted by both faculty members and students. Hence we tried to discriminate instructors' performance from the students point of view.

As a result of the research, we found out that the most powerful items that discriminate effective instructors from ineffective instructors are: "*satisfaction with the way course is conducted*", "*exams being a good measuring devise*", "*instructors being competent in his/her field*", " *instructors not lecturing in a logical sequence* (reverse item)","*finding the course very useful*", "*attending the course regularly*", "*instructors lecturing using real life examples*", and "*exam time being sufficient*".

These criteria are the result of a model with 81.5 % hit ratio, indicating a high predictive accuracy. For that reason we can advice instructors to consider these items in order to improve their teaching performance.

This study shows that if an instructor refines his way of conducting a course, uses real life examples, lectures in a logical sequence that can be followed by the students, cultivates his/her competence and prepares good measuring exam with sufficient time he /she will increase the willingness of students to take another course from him/her.

These are the results of students who find the course useful for their education life and attend the course regularly. Undoubtedly some students seem confused about the purpose and value of the ratings. Some fill them out as quickly as possible believing they will not make any difference in course content, and instructor teaching style, others do not fill it at all. To increase the quality of teaching, students should be convinced about the importance of their feedbacks so that student ratings will reflect what effective teaching is supposed to be.

Different dimension of teaching effectiveness are being measured in universities, and there is still some way to go in relation to conceptualizing the dimension of effective teaching and the development of measurement instruments. Sociological and psychological elements may affect students' choice of an effective instructor as well. For that reason our questionnaire can be improved further including other elements in the future.

This study should be repeated in several faculties and universities so that the generalizability of the results can be discussed. As it is the results can provide information for other faculties and universities. It is hoped the study can make a contribution to existing teaching evaluation literature.

REFERENCES

[1] JOHNSON, P., CASSELl, C., CLOSE, P., & DUBERLEY, J. (2001). **Performance Evaluation and Control Supporting Organizational Change. Management Decision.** Vol 39 No 10 pp 841-850.

[2] ARNOLD, H.J. & Feldman, D.C. (1986). **Organizational Behavior.** Singapore, McGraw Hill International Editions

[3] AAMODT, M.G. (1996). **Applied industrial/organizational psychology.** Brooks/Cole Publishing Company . Pasific Grove. CA.

[4] AHMADI, M., HELMS, M. M. & RAISZADEH, F. (2001). Business students' perceptions of faculty evaluations. **The International Journal of Educational Management.** Vol 15 No 1 pp 12-22.

[5] EMERY, C.R., KRAMER, T.R., & TIAN, R.G. (2003). Return to Academic Standards: a Critique of Student Evaluations of Teaching Effectiveness. **Quality Assurance in Education.** Vol 11 No 1 pp 37-46.

[6] RADMACHER, S.A. & MARTIN, D.J. (2001). Identifying Significant Predictors of Student Evaluations of Faculty Through Hierarchical Regression Analysis. **Journal of Psychology**. Vol 135 Issue 3 pp 259-269.

[7] CRUMBLEY, D.L. & FLIEDNER E. (2002). Accounting administrators' perceptions of student evaluation of teaching (SET) Information. **Quality Assurance in Education**. Vol 10 No 4 pp 213-222.

[8] LIAW, S. & GOH K. (2003). Evidence and Control of Biases ın Student Evaluation of Teaching. **The International Journal of Educational Management.** Vol 17 No 1 pp 37-43.

[9] COBURN, L. (1984). Student Evaluation of Teacher Performance. **ERIC/TME Update Series.** http://ericae.net/edo/ED289887.htm (Date 12.04.2003)

[10] HOBSON, S.M. & TALBOT,D.M. (2001). Understanding Student Evalutions **College Teaching.** Vol 49. Issue 1. pp 26-32.

[11] SOJKA, J., GUPTA, A.K. & Deeter-Schmelz, D.R. (2002). Student and Faculty Perceptions of Student Evaluations of teaching: A Study of similarities and Differences. **College Teaching**. Vol 20 Issue 2 pp 44-49.

[12] TIMPSON, W.W. & DESLEY, A. (1997). Rethinking Student Evaluations and The Improvement of Teaching: Instruments for Change at the University of Queensland. **Studies in Higher Education.** Vol 22 Issue 1 pp 55-66.

[13] KOH, H.C. & TAN, T. M. (1997) Empirical Investigation of the factors Affecting SET Results. **The International Journal of Educational Management.** Vol 11 No 4 pp 170-178.

[14] McKINNEY, K.(1997). What Do Students Ratings Mean? The National Teaching and Learning Forum. Vol 7 No 1 pp 1-4.

[15] WHITWORTH, J.E., PRICE, B.A. & RANDALL, C.H. (2002). Factors that Affect College of Business students Opinion of Teaching and learning. **Journal of Eduction for Business**. Vol. 77. Issue 5 pp 282-289.

[16] TC YÜKSEK ÖĞRETİM KURULU (2000). **Öğrencilerin Eğitimi Değerlendirmesi**. Ankara.: Ankara Üniversitesi Basım Evi

[17] HAIR, Jr. J.F., ANDERSON, E.R., TATHAM, R.L., & BLACK, C.W. (1998). **Multivariate Data Analysis** Upper Saddle River, NJ: Prentice Hall Inc.

[18] LATTIN, J.; CARROLL, D.; & GREEN, P.E. (2003) **Analyzing Multivariate Data.** Duxbury Applied Series. Pasific Grove, CA: Thomson Learning Inc.

[19] SHARMA, S. (1996). **Applied Multivariate Statistical Analysis.** New York, NY: John Wiley & Sons, Inc.

[20] NORUSIS, M.J. (1992). **SPSS for Windows: Advanced Statistics**. Chicago, IL: SPSS Inc.

[21] KLECKA, R.W. (1980). **Discriminant Analysis.** Newbury Park, California.: Sage Publications, Inc.

[22] BAYRAM, N.(2002). Diskriminant analizi: Akademisyenler üzerine bir uygulama. **Öneri**. Cilt 5. Sayı 17. s.219-229

[23] GREEN, S.B., SALKIND, N.J. & AKEY, M.T. (2000). **Using SPSS for Windows: Analysing and Understanding Data**. Upper Saddle River, NJ: Prentice Hall Inc.

[24] MENARD, S. (1995). **Applied Logistic Regression Analysis**. Thousand Oaks, California.: Sage Publications, Inc

[25] HOSMER, D.W. & LEMESHOW, S. (1989). **Applied Logistic Regression.** New York, NY: John Wiley & Sons, Inc.

Serra YURTKORU

Marmara Üniversitesi, İ.İ.B.F., İngilizce İşletme Bölümü Göztepe Kampüsü 34722 Kuyubaşı, Istanbul

0216 345 86 29

syurtkoru@marmara.edu.tr

E. Serra Yurtkoru has a Ph.D. of Business Administration at Marmara University Social Sciences Institute. She is Lecturer of Quantitative Methods at Marmara University. Her research areas are applied statistics, research methods, organizational change, leadership, consumer behavior, statistical process control, datamining

Beril SİPAHİ

Marmara Üniversitesi, İ.İ.B.F., İngilizce İşletme Bölümü Göztepe Kampüsü 34722 Kuyubaşı, Istanbul

0216 345 86 29

sipahi@marmara.edu.tr

Beril Sipahi has a Ph.D. of Business Administration at Marmara University Social Sciences Institute. She is Lecturer of Quantitative Methods at Marmara University. Her research areas are applied statistics, research methods, datamining, CRM, leadership, consumer behavior