DOCUMENT RESUME

ED 435 750 TM 030 402

AUTHOR	Zietz, Joachim; Cochran, Howard H., Jr.; Wilson, Mark L.
TITLE	Evaluating Instructional Effectiveness.
PUB DATE	1999-11-00
NOTE	17p.; Annual Meeting of the Mid-South Educational Research Association (Point Clear, AL, November 16-19, 1999).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*College Faculty; Economics; Evaluation Methods; Higher Education; *Instructional Effectiveness; *Outcomes of Education; *Student Evaluation of Teacher Performance; *Teacher Evaluation
IDENTIFIERS	Alternative Assessment

ABSTRACT

Instructor rankings for three alternative teaching performance measures were compared. The measures were an outcome measure (in this case, a test of economics knowledge), student evaluations of teaching (SET), and a weighted average of the two. For each performance measure, two alternative approaches to ranking were considered. One simply compared an instructor's score to those of other instructors. The other, a new method, compared an instructor's score to the one that could be expected for the teacher's class. The methodology was applied to rank 49 classes of principles of economic classes taught at comprehensive universities. Significant differences were found in instructor rankings between the three evaluation scores and also between rankings based on the suggested methodology and the traditional approach comparing an instructor to others. The newly developed methodology can be adapted to any college or department. In a first step, the suggested ranking methodology could be implemented for teaching evaluations alone. In a second step, evaluators could develop knowledge-based outcome tests for each field and course. Then a teacher's overall assessment could be calculated as a weighted average of the two scores. (Contains 5 tables and 14 references.) (SLD)

November 1999

Evaluating Instructional Effectiveness

Joachim Zietz
Howard H. Cochran, Jr.
Mark L. Wilson

Prepared for Presentation at the
Mid-South Educational Research Association
Annual Meeting, November 16 - 19, 1999
Point Clear, Alabama

Joachim Zietz
Dept. of Economics and Finance
Middle Tennessee State University
Murfreesboro, TN 37132

Tel.: (615) 895 6027
Fax : (615) 895 3010
E-mail: jzietz@mtsu.edu

Howard H. Cochran, Jr.
Belmont University
1900 Belmont Boulevard
Nashville, TN 37212

(615) 460 6477
(615) 460 6487
cochranh@mail.belmont.edu

Mark L. Wilson
Univ. of Charleston
2300 MacCorkle Ave.
Charleston, WV 25304

(304) 357 4865
(304) 357 4715
mwilson@uchaswv.edu

Evaluating Instructional Effectiveness

ABSTRACT

An instructor's teaching effectiveness for a particular class is assessed by comparing his/her actual evaluation score with the one that is predicted for the class rather than with the average score for all other instructors teaching the same class. The prediction is based on a random-effects regression model of the evaluation score on student, class, and teacher characteristics that are not under the control of the instructor. Teaching effectiveness is measured in this manner for a student evaluation score and a knowledge-based test score for 49 classes/instructors of principles of economics taught at comprehensive universities.

1. INTRODUCTION

Teaching excellence ranks high among universities' list of objectives. The most common approach to assessing teaching excellence consists of teaching evaluations by students toward the end of the term. They are in wide use across the U.S. (Seldin 1989). Student evaluations of teacher performance are likely to receive even more attention in the future as a new wave of outcomes assessment is sweeping across U.S. colleges and universities (McCoy et al. 1994).

Typically, the questionnaire filled out by students to evaluate instructors contains a question on the instructor's overall teaching performance relative to that of other teachers. University administrators tend to make heavy use of students' answers to this question because it seemingly cuts through the complexity of teaching evaluations: just one number says it all. It enters into formal assessments from a faculty's merit pay to decisions on tenure and promotion (White 1995).

An instructor's teaching can also be evaluated by identifying how much students have learnt during the semester. Compared to student evaluations, this is a considerably more involved task and, therefore, less popular. To examine a student's knowledge gain during the term, one would need to assess what students know at the end of the term compared to what they knew upon entering the class. Some efforts have been made in this respect in economics with the Test of Understanding in College Economics (TUCE), as developed by the Joint Council on Economic Education.

One may also think of assessing an instructor's teaching effectiveness by using a measure that combines both student evaluations and a test of knowledge gained, such as the TUCE. This could be a useful route if it is indeed true that student evaluations are not significantly correlated with student performance in the course, as suggested, for example, by Abrami et al. (1990) or Gramlich and Greenlee (1993).

Deciding between student evaluations and an outcomes measure, such as the TUCE, is one aspect of the larger issue of measuring teaching effectiveness. Generally, discussions stop at this point. The present study also considers another aspect of evaluating teaching effectiveness: the standard that is applied to evaluate an instructor's performance. Typically, an instructor's performance score is simply compared to the average score of all instructors. Although this method has the benefit of simplicity, it suffers from an apparent defect. Instructors teaching ill prepared students, large classes, mandatory courses, difficult material or at inconvenient times are likely to end up faring less well than instructors facing more favorable conditions in their courses. This type of problem can easily undermine the trust instructors have in the whole process of measuring teaching effectiveness. As such, it is clearly incompatible with an incentive structure that intends to promote increased productivity.[1] There is an alternative available: to compare an instructor's performance to the score that can reasonably be expected of him/her given the particular conditions that he/she has to deal with in his/her class. The key objective of this paper is to illustrate how that can be done.

The paper is organized as follows. The next section discusses the methodology. Subsequently, the methodology is illustrated with an application to a number of classes that were part of the TUCE III project. In particular, it is shown how instructor rankings can change rather significantly as one switches between alternative performance scores and/or between different methods of evaluating a particular performance score. The paper ends with a summary of its main points and a discussion of its implications for teaching assessment at universities.

2. METHODOLOGY AND DATA

The objective of this study is to compare instructor rankings for three alternative teaching performance measures: an outcomes measure (TUCE), student evaluations of teaching (SET), and a weighted average of the two. For each performance measure, two alternative approaches to ranking instructors are presented. The first approach simply compares an instructor's score to those of other instructors. The second approach compares an instructor's score to the one that can be expected for the instructor's class.

The first approach to comparing teaching scores is familiar and needs no explanation. The second approach represents this paper's key innovation. The idea is that instructors should be held responsible for a teaching evaluation or learning outcomes score only to the extent that he/she can influence it. The impact on the evaluation score of factors that are not under the control of the instructor should be ignored for evaluation purposes.

There are at least two broad groups of factors that could impact an instructor's evaluation score without being themselves under the control of the instructor: student characteristics, such as grade point average or student SAT scores, and course characteristics, such as class size. The literature is rich with evidence that the type of course as well as student characteristics influence student evaluations (e.g. Aigner and Thum 1986, Langbein 1994, Koon and Murray 1995). For example, required core curriculum courses tend to get lower evaluations than electives; the expectation of a good grade

---

1 Becker (1979) has shown theoretically that rewarding teachers for better teaching has little impact on the teaching outcome unless efforts are made simultaneously to improve the accuracy of measuring teaching ability.

improves evaluations; courses that demand more of a student's time outside of class have lower ratings; smaller classes increase student evaluations (Glass, McGaw, and Smith 1981). There is also ample evidence that student and course characteristics affect an outcomes measure such as the TUCE (e.g. Lopus and Maxwell 1995).

To circumvent the problem of attributing to instructors the consequences of conditions they are not responsible for, we compare an instructor's actual performance measure with the one that is predicted for his/her class. The prediction is based on a regression of actual performance on factors that are not under an instructor's control. There are numerous such factors. It is helpful to organize them around groups or categories that are reminiscent of the production function approach to economic education (Siegfried and Fels 1979): (i) student characteristics (SC), such as grade point average, previous knowledge the student has of the subject matter, SAT or ACT scores, a student's part-time or full-time status, (ii) class characteristics (CC), such as class size, meetings per week, time of day of class meetings, and (iii) certain teacher characteristics that are not under his/her control (TC), such as years teaching, years teaching the course under investigation, terminal degree, English as native language.

Three steps are involved in removing these factors from an instructor's performance score. First, a regression equation is specified that explains instructors' scores as a function of the above three groups of factors,

$$\text{Instructors' scores} = f(SC, CC, TC).$$

This equation is estimated for a data set that includes comparable classes, such as principles of economics at one institution or at several similar institutions. Second, the estimated regression is used for prediction purposes. For the particular class to be evaluated, class averages are calculated for all regressors and combined with the estimated coefficients to yield a predicted value of the instructor's score. Third, this predicted score is compared to the instructor's actual score. A teacher whose actual score exceeds the predicted one is considered to have performed above average or above expectations and vice versa. Instructor rankings can be based on the percentage difference between actual and predicted scores.

Since the variables that characterize an instructor's teaching style, effort, and talent are omitted from the regressions, the coefficients will be biased to the extent that the included variables are correlated with the omitted variables. Bias in the coefficients may have an impact on the rankings of instructors. A simple way around this potential problem would be to allow the intercept term to vary between classes/instructors. The intercept term would then incorporate all unobservable or unmeasured instructor effects. This points to either a fixed-effects (FE) or a random-effects (RE) model. A key advantage of the FE estimator is that its estimators remain consistent even if the instructor effects are correlated with the included regressors. Unfortunately, a fully specified FE-model is ruled out for the specified model by the type of data that are being analyzed: all class variables (CC) and all teacher variables (TC) are the same for all observations of a particular class. Hence, the choice between an FE- and an RE-model has to be made on the basis of the variable set SC alone.

Estimation is done for two performance scores: (i) an outcomes measure of learning (TUCE) and (ii) a measure of student evaluations (SET). The data come from the TUCE III data base (Saunders, 1991 and Saunders et al., 1991). The data set was assembled during the norming of the test in the fall and spring terms of the 1989-90 academic year. Of the 9,768 observations available, one observation per student, the study only utilizes those observations that relate to public comprehensive universities. The purpose is to reduce the amount of neglected heterogeneity. This reduced data set includes 49 different classes and 848 observations. The definitions of the variables are contained in Table 1.

3. RESULTS

In a first step, SET and TUCE are regressed only on student characteristics (SC) using both OLS, FE and RE estimators. The top part of Table 2 summarizes the essential test statistics. Incorporating cross-section effects of either the fixed or random type raises the R^2 appreciably. Both OLS regressions are rejected relative to the FE- and the RE-model alternatives at any common level of statistical significance. A Hausman test of the RE versus the FE-model is unable to reject the random-effects model in either case. Combining these results suggests that the RE-model is the preferred choice for the reduced data set, incorporating only SC.

The lower part of Table 2 provides the relevant statistics for the regressions on the expanded data set, incorporating CC and TC in addition to SC. R^2 levels go up for all estimated models. Since it is impossible to estimate the FE-model for the expanded data set, one cannot conduct a Hausman test for the RE-model. The ordinary least squares model is again rejected at any common level of statistical significance relative to the RE-model. Based on the results of Table 2, only the RE-models can be considered viable.

Table 3 reports for each class the percentage difference between actual TUCE and SET scores and both the sample average values and the predicted values from the two RE-models. The 49 classes are ranked according to the percentage by which the actual TUCE score for a class/instructor differs from the sample average TUCE score for the 49 classes/instructors. The top ranking class/instructor is listed first. Table 3 reveals that above average performance according to the TUCE are not necessarily associated with above average performance according to the SET. TUCE and SET do not appear to be closely correlated. This applies to the ranking based on average scores as well as to those that utilize predicted values from the RE-models. The low correlation of rankings based on TUCE and SET scores is underscored by the Spearman rank correlation coefficients reported in Table 4. For all three methods (average and both RE-models) the rank correlation is below 0.07 or even slightly negative. All three correlation coefficients are statistically insignificant from zero according to the z-scores provided in parenthesis.

Based on the rank correlation coefficients in Table 4, the RE(SC) model rankings are much closer to those based on averages than the rankings from the RE-model that is estimated for all available variables. The correlation for SET scores is particularly close (0.993) between the method based on averages and the one using the reduced variable set for prediction purposes (RE-SC). This close association suggests that much of the explanation of SET scores comes from the cross-section term v_i rather than from the variables in SC. This is confirmed by the low R^2 (0.026) for the corresponding OLS regression reported in the upper part of Table 2. The OLS-equation for SET explains somewhat more for the complete set of variables ($R^2 = 0.119$) and, as a consequence, there is more of a difference in rankings between this RE-model and the method based on averages (rank correlation = 0.775). The rankings based on equation predictions (RE-models) tend to have lower correlations with those based on averages for the TUCE than for the SET. Again, this can be explained by the higher R^2 for the TUCE equations than for the SET equations (Table 2). The higher the R^2 for the prediction equation, the lower is in all likelihood the correlation of rankings with the traditional method based on averages.

Table 3 reveals that striking changes in class/instructor performance evaluation can result as one moves from the traditional method based on averages to the suggested one based on predicted values. Let us assume first that the TUCE score is used for evaluation purposes. Then classes 7, 8, and 13, for example, would receive high marks since the actual TUCE score is considerably above the average TUCE score. Using the prediction method based on the RE model with a complete set of variables suggests, by contrast, that the same classes performed below expectations. In terms of instructor evaluation, the second method would indicate that the instructor has not done the job that one could have expected of him/her given the favorable conditions of the class (good students etc.). There are also numerous classes that show bad TUCE results based on the traditional method of averages but better than expected scores based on the prediction equation with the full set of variables. This applies, for example, to classes 22, 23, 35, and 47.

Similar discrepancies between performance scores based on averages and based on prediction equations result if one assumes that performance is measured by SET scores rather than TUCE scores. For example, classes 3, 15, 29, 34, and 48 show above average performance based on the traditional method of averages but below expectation scores based on the prediction equation (RE - SC, CC, TC). Classes 13, 17, and 46 are examples for low performance evaluations based on the traditional method but good evaluations based on the method suggested in this paper.

Table 5 ranks all 49 classes/instructors for each of the three methods examined in this paper. Three columns are provided for each method. The first column ranks classes based on the TUCE score of Table 3. The SET scores receive zero weight. Hence, the very first column of Table 5 repeats the ranking in the first results column of Table 3. The second column presents rankings if the TUCE and SET scores of Table 3 receive equal weight. The third column ranks classes by the SET score of Table 3, with the TUCE score receiving zero weight. Table 5 illustrates the rank correlation results presented in Table 4. It shows the large ranking differences between TUCE and SET regardless of the method used. It also clearly reveals the very close correlation of rankings based on SET scores for the traditional method based on averages and the prediction method based on the short list of variables (RE-SC). Table 5 shows that the class ranked 47th on the basis of its TUCE score by the traditional method of averages, ranks 7th for the RE (SC, CC, TC) method. Similarly, the eighth ranked class for the method of averages moves to rank 37 for the RE (SC, CC, TC) method.

4. SUMMARY AND CONCLUSIONS

The study has introduced a new method of ranking instructors' teaching effectiveness. It consists of comparing an instructor's actual evaluation score with the one that is predicted for him/her rather than with the average score for all other teachers. The prediction is based on a random-effects regression of the evaluation score on student, class, and teacher characteristics that are not under the control of the instructor.

The methodology was applied to rank 49 classes of principles of economics taught at comprehensive universities. Two basic evaluation scores were used: a knowledge-based outcomes test (TUCE) and student evaluations. Rankings were derived from these two for a third assessment score, one that puts equal weight on the knowledge-based outcomes test score and the teaching evaluation score. Significant differences were found in instructor rankings between the three evaluation scores and also between rankings based on the suggested methodology and the traditional approach of comparing an instructor's score to the corresponding average score for all instructors.

The methodology adopted in this paper can be adapted to any university, college, or department to make teaching assessment more meaningful. Additional variables that may be thought of as being important in determining an outcomes test or student evaluations but cannot be controlled by the instructor can be easily included in the model. Among such additional determinants that one may want to include are time-of-day the class is given, a variable that identifies whether a class is required for a student, or a variable that identifies the student as full-time or part-time. What variables are ultimately included in the prediction equation depends on the available data and the interests that prevail at a particular institution.

It is apparent that there is no need to rely on the TUCE as an outcomes measure. Any other outcomes test will also do. The only requirement for the suggested procedure is that one has data on a sufficient number of different classes/instructors before the prediction regression is run. This should not pose a problem in larger departments where a dozen or more parallel sections are taught for the same course. Nothing in the methodology limits its applicability to economics classes either. It is also not confined to principles classes, although it is more likely that a large number of concurrent sections are taught for principles than for other classes.

If one limits the analysis to student evaluations, the suggested methodology could be applied across the board to all undergraduate and graduate classes regardless of field. Rankings could be generated from a single regression equation for the whole institution. The only potential modification would be the inclusion of dummy variables to account for differences in graduate/undergraduate classes, by field of study or by college within the university. Compared to today's practice of dealing with student evaluations, this would make them significantly more comparable and, hence, more useful across classes and fields. With the given methodology it would be possible to rank all instructors of the university. Top teachers could be identified and rewarded every semester with great ease. There would also be little need to collect student answers on dozens of questions. One question on overall teacher performance would suffice as input for the above methodology. However, if one has student responses to other questions on the student evaluation form, it is of course possible to construct a prediction equation and to rank instructors for each question on the student questionnaire for which ranking makes sense.

Unfortunately, the results of this paper have confirmed earlier evidence by Abrami et al. (1990) and Gramlich and Greenlee (1993) that the correlation between instructor rankings based on a knowledge-based outcomes measure, such as the TUCE, and instructor rankings on the basis of student evaluations is very low. Good teaching evaluations do not necessarily mean that students learn a lot. If this result is taken seriously by university administrators, student evaluations cannot possibly be used in isolation to assess teaching effectiveness. They need to be supplemented with knowledge-based outcomes tests. This paper has offered a way to combine both in one measure of teaching effectiveness.

Looking toward the practical policy conclusions of this study, it seems that one could divide the path toward more useful teaching assessments into two steps. In a first step, the suggested methodology of ranking instructors by comparing actual to predicted scores could be implemented for teaching evaluations alone. This requires very few additional resources and can, therefore, be done almost immediately. In a second step, one would develop knowledge-based outcomes tests for each field and course. This would take significantly more time and resources and, for practical purposes, this step may be limited to principles classes or other large undergraduate or beginning graduate classes. Once the two steps are complete and all instructors teach at least one of the courses evaluated with an outcomes test, an instructor's overall teaching assessment would be calculated as a weighted average of the knowledge-based outcomes score for one of his/her classes and the SETs from all his/her classes taught during a given term.

REFERENCES

Abrami, P.C., S. d'Apollonia, and P.A. Cohen, "Validity of Student Ratings of Instruction: What We Know and What We Do Not." Journal of Educational Psychology 82 (June 1990): 219-231.

Aigner, D.J. and F.D. Thum, "On Student Evaluation of Teaching Ability," Journal of Economic Education 7 (Fall 1986): 243-265.

Becker, W.E., "Professorial Behavior Given a Stochastic Reward Structure," American EconomicReview 69 (December 1979): 1010-1017.

Glass, G.V., B. McGaw, and M.L. Smith, Meta-Analysis in Social Research. Beverly Hills, CA: Sage, 1981.

Gramlich, E.M. and G.A. Greenlee, "Measuring Teaching Performance," Journal of Economic Education 24 (Winter 1993): 3-13.

Koon, J. and H.G. Murray. "Using Multiple Outcomes to Validate Student Ratings of Overall Teacher Effectiveness." Journal of Higher Education 66 (Jan-Feb 1995): 61-82.

Langbein, L.I. "The Validity of Student Evaluations of Teaching." Political Science & Politics 27 (Sept 1994): 545-553.

Lopus, J.S. and N.L. Maxwell, "Teaching Tools: Should We Teach Microeconomic Principles Before Macroeconomic Principles?" Economic Inquiry 33 (1995): 336-350.

McCoy, J.P., Don Chamberlain, and Rob Seay, "The Status and Perceptions of University Outcomes Assessment in Economics," Journal of Economic Education 25 (Fall 1994): 358-366.

Saunders, P., "The Third Edition of the Test of Understanding in College Economics," American Economic Review, Papers and Proceedings 81 (May 1991): 32-37.

Saunders, P. et al. Examiner's Manual: Test of Understanding In College Economics. New York: Joint Council on Economic Education, 1991.

Seldin, P., "How Colleges Evaluate Professors." American Association for Higher Education Bulletin 41 (March 1989): 3-7.

Siegfried, J.J. and R. Fels, "Research on Teaching College Economics: A Survey." Journal of Economic Literature 17 (September 1979): 923-969.

White, L.J., "Efforts by Departments of Economics to Assess Teaching Effectiveness: Results of an Informal Survey." Journal of Economic Education 26 (Winter 1995): 81-85.

Table 1. Definition of Variables

Variable Group	Variable	Definition
Dependent Variables	TUCE	number of total correct answers on final test
	SET	student evaluation score
Student Characteristics (SC)	GEN	1 if gender is male
	GPA	cumulative grade point average
	PRE	score on TUCE pre-test
	SATACT	SAT or ACT score - converted to common scale
	RACE2-RACE4	race; RACE1 = white is base
	GPASAT	GPA * SATACT
Class Characteristics (CC)	CLAS1	1 for class size 31 to 40
	CLAS2	1 for class size 41 to 50
	CLAS3	1 for class size 51 to 75
	CLAS4	1 for class size 76 to 100
	CLAS5	1 for class size 101 to 200
	D2WK	1 for class with two weekly meetings
	D3WK	1 for class with three weekly meetings
	D4WK	1 for class with four weekly meetings
Teacher Characteristics (TC)	ENG	1 if instructor's native tongue is English
	PHD	1 if instructor holds doctorate
	YRSTCH	years instructor has been teaching
	YRSTCHCS	years instructor has been teaching course
	CLASTRM	number of classes taught by instructor per term
	CORSTRM	number of courses taught by instructor
	CLAST2	number of classes taught squared
	CORS2	number of courses taught squared

Notes: Class size is defined as the sum of the number of observations in each course.

Table 2. Least Squares, Fixed-Effects, and Random-Effects Models Compared

	SET			TUCE		
	OLS	FE	RE	OLS	FE	RE
SC variables only						
R^2	0.026	0.287	0.194	0.447	0.591	0.502
adj. R^2	0.010	0.232	0.131	0.441	0.559	0.464
OLS vs. FE	0.000			0.000		
RE vs. FE			0.482			0.240
OLS vs. RE (p-values)	0.000			0.000		
SC, CC, TC variables						
R^2	0.119	na	0.226	0.527	na	0.537
adj. R^2	0.092	na	0.148	0.513	na	0.491
OLS vs. RE (p-values)	0.000			0.000		

Notes: OLS stands for ordinary least squares without group effects; FE indicates a fixed-effects, RE a random-effects model. A low p-value, such as 0.000, means the null hypothesis is rejected.

BEST COPY AVAILABLE

Table 3. Actual Score as a Percentage of Average and Predicted Scores

Class No.	Percentage of Average Score		Percentage of Predicted Score			
			RE (SC)		RE (SC, CC, TC)	
	TUCE	SET	TUCE	SET	TUCE	SET
1	59.0	-13.8	32.6	-10.3	22.0	-10.2
2	55.1	8.4	25.1	12.3	20.9	21.3
3	51.2	9.7	13.5	10.6	6.8	-2.4
4	40.5	6.9	13.6	8.6	12.2	1.2
5	38.7	-4.0	29.0	-3.7	11.4	1.2
6	28.9	-14.5	23.2	-14.1	1.5	-3.8
7	28.3	1.8	14.5	2.6	-5.2	-0.5
8	27.9	8.4	19.3	9.2	-6.4	2.6
9	25.6	-10.8	5.6	-9.3	19.1	-6.1
10	25.4	13.1	16.6	13.5	14.4	19.7
11	22.1	3.2	9.6	3.4	5.5	-0.9
12	21.9	-19.0	1.7	-18.7	8.0	-10.1
13	20.6	-9.5	10.0	-8.4	-0.1	3.9
14	15.7	13.4	0.1	13.7	3.6	10.0
15	15.2	7.1	18.8	6.8	10.3	-1.8
16	12.2	12.8	13.8	13.0	13.7	7.9
17	3.9	-6.9	14.1	-4.8	19.9	3.7
18	3.1	5.1	5.5	10.1	-3.1	2.4
19	3.0	-9.9	0.6	-8.0	1.5	-3.3
20	1.6	-0.1	-2.6	1.2	-4.4	2.4
21	1.4	-0.0	-17.0	0.3	-5.6	5.4
22	0.0	-4.9	5.6	-4.0	12.2	-12.0
23	-1.9	6.8	-3.2	6.5	10.9	8.5
24	-2.8	-4.0	-3.6	1.9	2.2	1.5
25	-3.1	-35.1	-8.1	-33.0	-8.4	-25.4
26	-4.0	10.9	-4.8	11.3	-1.6	6.4
27	-4.4	-12.1	-0.3	-10.1	-4.6	-8.5
28	-4.4	13.1	-4.4	13.1	-6.8	14.9
29	-5.0	9.7	-2.8	9.9	-13.0	-1.5
30	-6.0	-3.5	-1.5	-3.6	-0.4	-6.8
31	-6.4	-24.6	-0.2	-23.4	-0.5	-23.8
32	-9.0	5.0	-14.3	5.4	-8.0	3.7
33	-9.4	-22.1	-4.4	-22.2	-3.4	-23.6
34	-9.5	5.3	-8.8	6.2	-6.0	-4.1
35	-9.5	-10.3	-10.1	-9.3	5.9	-3.9
36	-11.2	-11.0	-5.7	-10.9	-7.5	-9.8
37	-12.8	4.2	-13.2	4.3	-4.5	4.7
38	-13.9	10.6	-14.2	11.9	-8.5	2.8
39	-14.7	14.6	-13.0	17.7	-14.8	15.8
40	-15.1	3.5	-8.5	4.0	-0.7	2.4
41	-16.9	-4.6	-14.5	-3.8	-1.5	-2.0
42	-17.3	-1.4	-8.1	-1.6	-2.0	-3.4
43	-18.0	-12.7	-14.7	-10.7	-7.5	-5.8
44	-20.0	-21.6	-17.4	-21.4	-16.8	-9.3
45	-20.2	-9.5	-15.6	-9.0	-0.4	-4.3
46	-21.7	-6.7	-19.4	-5.7	-12.2	6.7
47	-21.8	0.4	11.5	2.3	13.5	-0.8
48	-25.7	6.8	-14.7	6.9	-7.4	-2.1
49	-27.2	17.4	-21.7	18.5	-20.3	10.9

Note: RE stands for random-effects model. (SC) indicates that only the variables in set SC of Table 1 were used for the prediction equation. (SC, CC, TC) means that all variables of Table 1 were used for prediction purposes. Classes are ranked according to performance as measured by the percentage difference between actual and sample average score (first results column).

Table 4. Spearman Rank Correlation Coefficients

	Average		RE (SC)		RE (SC, CC, TC)	
	TUCE	SET	TUCE	SET	TUCE	SET
Average						
TUCE		0.056 (0.38)	0.850 (5.89)		0.616 (4.27)	
SET				0.993 (6.88)		0.775 (5.37)
RE (SC)						
TUCE				0.067 (0.46)	0.722 (5.00)	
SET						0.782 (5.42)
RE (SC, CC, TC)						
TUCE						-0.016 (-0.11)

Notes: Numbers in parenthesis are z-scores. A value above 1.64 in absolute terms identifies statistical significance at better than the five percent level.

Table 5. Class/Instructor Rankings for Alternative Methods:
Average versus Prediction Equations and Alternative TUCE Weights

	Average			RE (SC)			RE (SC, CC, TC)		
TUCE Weight:	1	0.5	0	1	0.5	0	1	0.5	0
	1	2	49	1	2	49	1	2	2
	2	3	39	5	10	39	2	10	10
	3	4	14	2	8	14	17	17	39
	4	1	28	6	16	10	9	16	28
	5	10	10	8	15	28	10	23	49
	6	8	16	15	5	16	16	14	14
	7	5	26	10	3	2	47	4	23
	8	7	38	7	1	38	4	9	16
	9	14	29	17	4	26	22	47	46
	10	11	3	16	7	3	5	5	26
	11	16	8	4	18	18	23	1	21
	12	15	2	3	14	29	15	15	37
	13	9	15	47	47	8	12	28	13
	14	6	4	13	11	4	3	26	32
	15	13	23	11	17	48	35	11	17
	16	28	48	22	6	15	11	3	38
	17	18	34	9	28	23	14	13	8
	18	26	18	18	29	34	24	24	20
	19	23	32	12	26	32	19	35	40
	20	29	37	19	39	37	6	40	18
	21	12	40	14	23	40	13	39	24
	22	20	11	31	22	11	45	37	5
	23	21	7	27	13	7	30	22	4
	24	39	47	30	20	47	31	21	7
	25	17	21	20	24	24	40	18	47
	26	38	20	29	38	20	41	19	11
	27	32	42	23	34	21	26	20	29
	28	34	30	24	49	42	42	12	15
	29	22	24	33	9	30	18	6	41
	30	24	5	28	40	5	33	41	48
	31	19	41	26	30	41	20	8	3
	32	37	22	36	19	22	37	32	19
	33	30	46	42	48	17	27	45	42
	34	49	17	25	37	46	7	42	6
	35	40	13	40	32	19	21	46	35
	36	27	45	34	42	13	34	38	34
	37	42	19	35	27	45	8	7	45
	38	48	35	39	36	35	28	30	43
	39	35	9	37	21	9	48	49	9
	40	47	36	38	12	27	43	48	30
	41	41	27	32	41	1	36	34	27
	42	36	43	41	35	43	32	27	44
	43	46	1	48	31	36	25	43	36
	44	45	6	43	45	6	38	29	12
	45	43	12	45	46	12	46	36	1
	46	31	44	21	43	44	29	31	22
	47	33	33	44	33	33	39	44	33
	48	25	31	46	44	31	44	33	31
	49	44	25	49	25	25	49	25	25

Note: Classes are ranked from best to worst. The numbers refer to the class numbers identified in Table 3, which are identical to the ones in the first column of this table.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM030402

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Evaluating Instructional Effectiveness	
Author(s): Howard Cochran, Joachim Zietz, Mark Wilson	
Corporate Source:	Publication Date: 11-18-99

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☒	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here,→ please

Signature: Howard Cochran	Printed Name/Position/Title: Associate Professor Economics	
Organization/Address: Belmont University, 1900 Belmont Blvd, Nashville, TN 37212	Telephone: 615-460-6477	FAX: 615-460-6487
	E-Mail Address: cochranh@mail.belmont.edu	Date: 11-17-99

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com