DOCUMENT RESUME

ED 455 311 TM 033 171

AUTHOR	Chang, Te-Sheng
TITLE	An Application of Regression Models with Student Ratings in Determining Course Effectiveness.
PUB DATE	2000-04-00
NOTE	22p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Difficulty Level; Evaluation Methods; Foreign Countries; Higher Education; *Instructional Effectiveness; *Regression (Statistics); *Student Evaluation of Teacher Performance; *Undergraduate Students; Validity
IDENTIFIERS	Taiwan

ABSTRACT

The purpose of this study was to construct regression models that can identify sources of desired or undesired influences on student ratings. The approach uses course effectiveness findings developed with regression models to identify a possible solution to the two problems in using student ratings as a major component of course effectiveness: (1) variables other than teacher performance may contribute to ratings; and (2) students may tend to be generous in their ratings. The sample included 114 undergraduate courses from the Department of Elementary Education at National Hualien Teachers College, Taiwan, in 1999. The Student Ratings of Instruction form was used to measure students' perceptions of faculty performance. Five background variables are included in the developed regression equation. They are student enthusiasm, participation, expected grade, grading standard, and course difficulty. The results of this study indicate that 99.1% of the courses are rated above the middle of the entire raw-score scale. The T scores (adjusted scores), converted from the residual in regression models, are between 16.45 and 74.94. Twenty courses rated effective by the unadjusted score are classified as ineffective by the adjusted score. Eighteen courses judged ineffective through the unadjusted score are classified as effective through the adjusted score. The consistency of course-ranking classification is 66.7%. The correlation between unadjusted scores and adjusted scores in 0.447. The correlation between unadjusted course rankings and adjusted course rankings is 0.334. (Contains 37 references and 8 tables.) (SLD)

Running head: COURSE EFFECTIVENESS

An Application of Regression Models with Student Ratings in Determining Course effectiveness

Te-Sheng Chang
National Hualien Teachers College
Hualien, Taiwan, 970, R.O.C.
achang@sparc2.nhltc.edu.tw

This article was presented at the annual meeting of American Educational Research Association, held in New Orleans, LA, April 24-28, 2000.

Abstract

There are two problems in using student ratings as a major component of course effectiveness. First, variables other than teacher performance may inappropriately contribute to student ratings. Second, students may tend to be generous in their ratings. The purpose of this study is to construct regression models that can identify sources of desired or undesired influences on student ratings. Specifically, this study presents course effectiveness developed with regression models in identifying a possible solution that provides reasonable answers to these questions.

The sample included 114 undergraduate courses from Department of Elementary Education at National Hualien Teachers College in the spring semester 1999. The Student Ratings of Instruction (SRI) form was used to measure students' perceptions of faculty performance.

Five background variables are included in the final regression equation. They are student enthusiasm, participation, expected grade, grading standard, and course difficulty. The results of this study indicate that 99.1% of courses are rated above the middle of the entire raw-score scale. The T scores (adjusted scores), converted from the residual in regression models, are between 16.45 and 74.94. Twenty effective courses by the unadjusted score are classified as ineffective by the adjusted score. Eighteen ineffective courses by the unadjusted score are classified as effective by the adjusted score. The consistency of course- ranking classification is 66.7%. The correlation between unadjusted scores and adjusted scores is .447. The correlation between unadjusted course rankings and adjusted course rankings is .334.

Key words: Student Ratings of Instruction, Faculty Evaluation, Course Effectiveness

An Application of Regression Models with Student Ratings in Determining Course Effectiveness

Te-Sheng Chang
National Hualien Teachers College

Introduction

Background

Procedures for measuring faculty teaching performance as course effectiveness are different from schools to schools, however, student ratings of instruction are consistently considered in the process and are typically important elements of tenure and promotion decisions. Some realistic for student ratings of instruction to be a major component in teaching evaluation are as follows: (1) Students are an obvious and convenient choice for raters. (2) They have closely and recently observed a number of teachers. (3) They particularly know how students think and feel. (4) Students' frank reactions can be a beneficial aid in refining course structure and teaching styles. (5) Student ratings are more objective than many other approaches, such as administrator evaluation, peer evaluation, teacher self-evaluation, and classroom visitations (Arreola, 1995; Feldman,1997; Jirovec, Ramanathan, & Alvarez, 1998; Wachtel, 1998).

Because of the high correlation between quality teaching and high student achievement (Brown, 1977), it is reasonable that course effectiveness of faculty should be carefully monitored and explained. Student ratings of faculty course effectiveness are also used in dispensing merit and can create a competitive climate among faculty within colleges and departments. Since the emphases place on student ratings and the pressure for faculty in particular to be rated high, an examination of how student ratings reflect the quality of teaching or course effectiveness is important.

Despite the benefits and rationales offered to justify use of student ratings, many faculty members have hesitatingly warmed to the concept. The weakness of student ratings is their uncertainty in reflecting the quality of teaching (Jirovec, et al., 1998). According to literature, faculty's suggestion (from the school in which the researcher teaches), and researcher's experience, there are two problems in using student ratings as a major component of course effectiveness. First, factors other than teacher performance (e.g., course difficulty, grading standard, student motivation, etc.) may inappropriately contribute to student ratings (Chang, 1997; Feldman, 1993; Marsh & Roche, 1997; Wachtel, 1998). Second, students may tend to be generous in their ratings. Without dealing with these two problems, the faculty/course evaluation goals and student ratings program's demands can not be met.

Purposes

Due to the two problems described above, the purpose of this study is to construct regression models that can identify sources of desired or undesired influences on student ratings. Specifically, this study presents course effectiveness developed with regression models in identifying a possible solution that provides reasonable answers to these areas in question.

Perspectives

The growing call for the use of student ratings of instruction in the definition and selection of effective classes (or faculty) has led to a new set of methodological problems in identifying these classes. Faculty in the colleges can rarely select the students who attend. For example, a class or an instructor may have a high proportion of students with high concentration for the course; another class or instructor a mix of high and low concentration students, with varying percentages on concentration (students take the course in their major or minor). An underlying assumption behind this study is that "student ratings are biased to the extent that they are influenced by variables unrelated to course effectiveness". (Marsh, 1984, pp.733-734). Thus, it is unfair to compare the evaluations from high student concentration courses to low student concentration courses, although the evaluations themselves are not necessarily biased. The problems, then, come from the necessity of making fair comparisons among the classes in determining effectiveness.

Fairness questions can be corrected by analytic techniques. The most satisfactory approach to comparing course effectiveness with these different characteristics of courses is to examine course outcomes (student ratings) in relation to expected or predicted outcomes. This is becoming known as the "value-added" approach to determining effectiveness (Mendro, Webster, Bembry, & Orsak, 1995). In essence, value-added methodologies determine a predicted outcome on student ratings for a course with a given set of background characteristics. Course effectiveness is determined by how much the course exceeded or fell below the predicted value of the student ratings at the course.

Variables Thought to Influence Student Ratings

What variables may not be directly related to course effectiveness but affect the results of student ratings? A wealth of research exists in the area of student ratings, ranging from analyses of validity and reliability to studies parceling effects related to course, student, and teacher characteristics. This section provides a simple overview of the findings related to the variables, which could conceivably exert an influence on student ratings scores.

Course Characteristics

Researchers reported that teachers of elective or non-required courses received higher ratings than teachers of required courses; a small to moderate positive relationship was found between course electivity and evaluation scores (Marsh & Roche, 1997; Scherr & Scherr, 1990). This might be due to lower prior subject interest in required versus non-required courses. Most studies found that higher level courses tend to receive higher ratings (Chang, 1997; Marsh, 1987). Chang explained that students in high level courses might have more learning enthusiasm toward courses than those in lower level courses. Feldman (1978) reported that the association between course level and ratings is decreased when other background variables such as class size, expected grade, and electivity are controlled.

Greenwald and Gillmore (1998) reported that the introduction of mandatory student ratings led faculty to reduce course workloads and to make examinations easy in order to receive higher evaluation scores. They examined student ratings of hundreds of courses at University of Washington and found that professors who are easy graders receive better evaluations than do professors who are tougher. Marsh (1980) and Franklin, Thell, and Ludlow (1991), on the other hand, found a positive effect of course difficulty where more difficult courses were rated higher than less difficult courses. Wachtel (1998) argued that course level and student age might be confounding factors in more difficult courses.

Studies examining class size have arrived at various conclusions. Most researchers found that smaller classes tend to receive higher ratings (McKeachie, 1990). Marsh and Dunkin (1992) argued that the class size effect is specific to certain dimensions of effective teaching performance, namely group interaction and instructional rapport. Another hypothesis was that the relationship between class size and student ratings is a U-shaped or curvilinear relationship, with small and large classes receiving higher ratings than medium-sized ones (Feldman, 1984). Some explanations which have been offered for this relationship included: departments may assign known superior teachers to large lecture classes or superior teachers may attract more students to their classes by virtue of their reputation (Wachtel, 1998).

Student Characteristics

Evidence suggested that students with greater interest in the subject area prior to the course tend to give more favorable teacher ratings (Prave & Bairl, 1993). Marsh and Dunkin (1992) asserted that the influence of prior interest on student ratings does not constitute a bias. They admitted that when ratings are used for summative purpose, the influence of student interest toward a subject can be a source of unfairness in that, but it is a function of the course and not the teacher.

The effect of a student's expected grade in a course on the student ratings has been one of the most controversial topics. Numerous authors argued in favor of the leniency hypothesis (Koshland, 1991; Nimmer & Stone, 1991) and against it (Marsh, 1987; Theall & Franklin, 1991). However, at this time, the consensus was definitely that there is a moderate positive correlation between expected grade and student ratings (Braskamp & Ory, 1994; Marsh, 1987; Marsh & Dunkin, 1992). The controversy concerned the interpretation of this association. Chacko (1983) showed that more strict grading standards led students to rate the instructor lower even on components of instruction unrelated to grading fairness, such as humor, self-reliance, and attitude toward students. Marsh (1987) gave three plausible interpretations: the leniency hypothesis, the validity hypothesis, and the student characteristics hypothesis. In the leniency hypothesis, instructors with more lenient grading standards receive more favorable ratings. In the validity hypothesis, more effective instructors cause students to work harder, learn more and earn better grade. In the student characteristics hypothesis, pre-existing student characteristics such as prior subject interest affect both course effectiveness and student ratings.

The effect of student gender on student ratings is another controversial topic. Many studies reported that there was essentially no difference in ratings by male and female students, but a few have also come to a different conclusion (Watchel, 1998).

Tatro (1995), for example, found that female students gave higher ratings than males. However, Koushki and Kuhn (1982) found the opposite results. In addition, some studies reported a tendency for student to rate same-sex instructors slightly higher than opposite-sex instructors (Centra, 1993; Feldman, 1993).

Teacher Characteristics

Research typically indicated a positive effect of teacher rank on student ratings but a negative effect for age of the faculty member and years of teaching on ratings (Feldman, 1983). Feldman noted that while higher faculty rank is typically associated with higher overall ratings, the relationship can disappear or reverse when particular dimensions of teaching are examined. Discussion of the effect of teacher gender on student ratings appeared to be quite varied. In a two-part meta-analysis, Feldman (1992, 1993) reviewed existing research on student ratings of male and female teachers in both the laboratory and the classroom setting. In his review of laboratory studies, Feldman (1992) reported that the majority of studies reviewed showed no difference in the global evaluations of male and female teachers. In the minority of studies, in which difference was found, male instructors received higher overall ratings than females. Subsequently, in his review of classroom studies, Feldman (1993) again reported that the majority of studies reported no significant differences between the genders.

Grading standard perhaps generates the most suspicion about the validity of student ratings. Bridgeman (1986) and Owie (1985) compared summary evaluation scores of three groups, those receiving grades worse than expected, same as expected, and better than expected. Both of them found significant differences among the groups. The lowest evaluations came from the negative-discrepancy group; the highest came from the zero-discrepancy group for Bridgeman and the positive-discrepancy group for Owie.

Greenwald and Gillmore (1997) have given five theories of the positive relationship between grades and student ratings: (1) Course effectiveness influences both grades and ratings. (2) Students' general academic motivation influences both grades and ratings. (3) Students' course-specific motivation influences both grades and ratings. (4) Students infer course quality and own ability from received grades. (5) Students give high ratings in appreciation for lenient grading. They interpreted that the existence of this grades-ratings correlation prompts a suspicion that ratings can be increased by the strategy of increasing grades, but by no means does it demand that conclusion. The first three theories explain the grades-ratings correlation by assuming that a third variable influences both grades and ratings. By appealing of a causal influence of grades on ratings. The remaining two theories do assume that grades have a causal influence on ratings (Greenwald & Gillmore, 1997).

Besides the background characteristics this study have discussed, Marsh and Roche (1997) have summarized research studies on the relationship between students' ratings and background characteristics. Table 1 presents their study.

Table 1. *Overview of Relationship Found between Students' Ratings and Background Characteristics by Marsh & Roche (1997)*

Background characteristic	Summary of findings
Prior subject interest	Classes with higher interest rate classes more favorably, although it is not always clear if interest existed before the start of the course or was generated by the course or the instructor.
Expected grade-actual grade	Class-average grades area correlated with class-average students' evaluations of teaching, but the interpretation depends on whether higher grades represent grading leniency, superior learning, or preexisting differences.
Reason for taking a course	Elective courses and those with a higher percentage of students taking the course for general interest tend to be rated higher.
Workload-difficulty	Harder, more difficult courses requiring more effort and time are rated somewhat more favorably.
Class size	Mixed findings but most studies show smaller classes are rated somewhat more favorable, although some find curvilinear relationships where large classes also are rated favorably.
Level of course or year in school	Graduate-level courses are rated somewhat more favorable; weak, inconsistent findings suggest upper division courses are rated higher than lower division courses.
Instructor's rank	Mixed findings but little or no effect.
Sex of instructor or student	Mixed findings bur little or no effect.
Academic discipline	Weak tendency for higher ratings in humanities and lower ratings in sciences, but too few studies to be clear.
Purpose of ratings	Somewhat higher ratings if ratings are known to be used for tenure-promotion decisions.
Administrative conditions	Somewhat higher if ratings are not anonymous and the instructor is present when ratings are being completed.
Students' personality	Mixed findings but apparently little effect, particularly because different personality types may appear in somewhat similar numbers in different classes.

Note. Particularly for the more widely studies characteristics, some studies have found little or no relation or even results opposite to those reported here. The size, or even the direction, of relations may vary considerably, depending on the particular component of students' ratings that is being considered. Few studies have found any of these characteristics to be correlated more than .30 with class-average students' ratings, and most relations are much smaller.

Based on the past research studies, the directions of the relationships between student ratings and certain background characteristics are mixed and magnitude of the relationships tend to be small. Two points must be noted. First, the size and direction of the relationship between background characteristics and student ratings seem to lie in the situation and condition in which the former studies were conducted. Second, although the effects of background characteristics on student ratings are mixed, they need to be taken into consideration when student ratings are applied for the determination of course effectiveness.

Method

Sample

The data for this investigation came from Department of Elementary Education at National Hualien Teachers College in Taiwan. Student ratings of department faculty were collected in the spring semester of the 1998-1999 academic year. Evaluations on which students failed to respond to questions that are key variables in the model were eliminated. The final analytic sample included 114 undergraduate courses with 23(20.2%) freshman classes, 33(28.9%) sophomore classes, 37(37.7%) junior classes, and 15(13.2%) senior classes. It was possible that one instructor was rated by several courses and that one student contributed several ratings to the database. Given the sample size, it was expected that the effects of these repeated observations would be negligible.

Instrument

The Student Ratings of Instruction form (SRI) developed by the faculty evaluation committee was used to measure students' perceptions of teacher appeal and course effectiveness during the last two weeks of classes. The rating form was composed of 13 questions rated on a 5-point Likert scale ranging from strongly agree (5-point) to strongly disagree (1-point). The average of these 13 items was considered as the overall rating score for an instructor's course effectiveness within a course.

Principal components analysis was applied to examine the construct validity of the instrument. Factor loadings were large, between .706 and .939. There was only one eigenvalue greater than 1 (9.14), which indicated the items were pure indicators for their own factor. This overall factor accounted for 76% of the total variance. The α coefficient of internal consistency reliability was .969, which confirmed that the questionnaire was a reliable instrument.

Background Variables

Information about course, class, student, and instructor characteristics was obtained on 13 different variables.

(1) course difficulty: Student perceptions of the relative difficulty required by the course. An evaluation item score ranges from 1 to 5, 1:very easy; 3:medium; 5:very difficult.

(2) course level: There are four levels for the course division, 1 for freshman, 2 for sophomore, 3 for junior, and 4 for senior division.

(3) type of course requirement: Courses are classified into either required (assigned as 0) or elective courses (assigned as 1) according to their status in the curriculum.

(4) concentration: Students take the course in their major (assigned as 1) or not (assigned as 0) for example, the mathematics students in a mathematics course.

(5) class size: The number of students are enrolled in the class.

(6) enthusiasm toward the subject: Level of student enthusiasm for the subject or course. An evaluation item score ranges from 1 to 5, 1:very low; 3:medium; 5:very high.

(7) student participation: Frequency of student participation into the class for the semester. An evaluation item score ranges from 1 to 5, 1:seldom; 3:medium; 5:always.

(8) expected grade: The final grade students expected the instructor would give to them. An evaluation item score ranges from 1 to 5, 1:below 60; 2:60 to 69; 3:70 to 79; 4:80 to 89; 5: above 90.

(9) teacher gender: 1:male instructor; 0:female instructor.

(10) teacher rank: 1:full professor; 2:associate professor and assistant professor; 3: lecturer.

(11) teacher age: Instructor age was computed as with the formula of the year instructor born subtracted from 1998. (e.g., A teacher was born in 1961, he would be 36 years old in this study).

(12) teacher degree: 1: bachelor, 2: master, 3: doctor.

(13) grading standard: The discrepancy between student-expected grade and the grade students thought their teachers would give to them. Positive discrepancy means grading standard is strict; while negative discrepancy means teacher grading standard is lenient. That is, the higher the discrepancy, the stricter the grading standard.

Design and Data Analysis

All analyses were performed on class-average responses for the sample. Thirteen background characteristics obtained from the survey and school database were course difficulty, course level, electivity, concentration, class size, student enthusiasm toward the subject, participation, expected grade, teacher gender, rank, age, degree, and grading standard.

The zero-order correlation, semi-partial correlation, and stepwise regression were used to determine which of the background variables made the largest contribution and to develop the best linear regression models. The linear regression model developed for the course effectiveness controlled for the "determined" variables affecting student ratings. The student rating score was regressed on all the "determined" variables. The effects of the "determined" variables were moved by subtracting the original score from the regression estimate of each score (predicted score). The residual between the unadjusted score and the predicted score was converted to T score. The T score was referred to as adjusted course effectiveness score. The courses scoring at or above the median of the unadjusted score were classified as unadjusted effective courses. Those scoring below the median of the unadjusted score were classified as unadjusted ineffective courses. Similarly, the courses scoring at or above the median of the adjusted score were classified as adjusted effective courses. Those scoring below the median of the adjusted score were classified as adjusted ineffective courses.

The correlation between the unadjusted score (the raw score) and the adjusted score (the T scores converted from the residual) was assessed by Pearson product-moment correlation coefficient. Similarly, the correlation between the unadjusted rank and the adjusted rank was computed by Spearman rank correlation coefficient.

Results

Table 2 shows the zero-order correlation and semi-partial correlation coefficients between each of the 13 background variables and the evaluation score. Five of the 13 zero-order correlation coefficients are statistically significant ($p < .05$) and account for at least 5 percent of the variance in the evaluation score ($r \geq .23$). They are student enthusiasm, participation, expected grade, course difficulty, and class size. The first three variables are positively correlated with the evaluation score, while course difficulty and class size are negatively correlated with the evaluation score. Although 5 correlations between background variables and the evaluation score account for 5 percent of the variance, only the semi-partial correlation coefficients of student enthusiasm, participation and the evaluation score are greater than .10. That is, only student enthusiasm and participation uniquely explain at least 1% of the variance in the evaluation score.

Table 2. *Correlation and Semi-partial Correlation between 13 Background Variables and Student Ratings Score (N=114)*

Background variables	*r*	*sr*
Course		
Difficulty	-.780**	-.026
Level	.120	.016
Electivity	-.044	-.053
Concentration	-.115	.010
Size	-.239*	-.025
Student		
Enthusiasm	.870**	.240*
Participation	.832**	.113
Expected grade	.573*	.022
Instructor		
Gender	-.131	-.013
Rank	-.082	.017
Age	-.131	.045
Degree	.040	-.027
Grading standard	-.093	-.085

* $p < .05$ ** $p < .01$ *r*: Pearson product-moment correlation. *sr*: semipartial correlation.

Table 3 presents the summary of stepwise regression analysis for background variables predicting the evaluation score. Four background variables are maintained in the final regression models. They are student enthusiasm, participation, teacher grading standard, and teacher age. The percentage of variance explained by this final combination of background variables is 80.2%. The attention is paid to the variables only if the change in total variance accounted for from the step is greater than .01(1%). Therefore, teacher age is not maintained in the regression model.

Table 3. *Summary of Stepwise Regression Analysis for Background Variables Predicting the evaluation score (N = 114)*

Step	Variable	*b*	*SE b*	*β*	*R*	R^2	ΔR^2	*F*
1	*a* = 1.489							
	Student enthusiasm	.626	.033	.870	.870	.757	.757	349.403***
2	*a* = 1.428							
	Student enthusiasm	.426	.061	.592	.886	.785	.028	202.829***
	Participation	.211	.056	.324				
3	*a* = 1.425							
	Student enthusiasm	.453	.062	.629	.891	.795	.010	141.229***
	Participation	.186	.056	.286				
	Grading standard	-.009	.044	-.096				
4	*a* =1.223							
	Student enthusiasm	.455	.061	.633	.896	.802	.008	110.559***
	Participation	.199	.055	.305				
	Grading standard	-.113	.044	-.112				
	Teacher age	.003	.001	.096				

* $p < .05$ ** $p < .01$; *a:* intercept; ΔR^2 : the increment of R^2. The values which are underlined indicate the increment of R^2 greater than 1%.

Based on the literature and the results of this study (Tables 2 and 3), there may be five background variables which have most contribution to student ratings in terms of practical and statistical significance. They are student enthusiasm, participation, expected grade, teacher grading standard, and course difficulty. The final regression model is established with these five variables. Table 4 shows the summary of final regression analysis. The multiple regression using the five predictors simultaneously yields $R^2 = .800$. Namely, the regression explains 80% of the variance in the evaluation score. The analysis yields the following equation to compute a score that is adjusted for effects of the five predictors. Residual = unadjusted score - [1.581 + .408 (student enthusiasm) + .146 (participation) + .074 (expected grade) -.144 (grading standard) -.061 (course difficulty)]. The residual is converted to T score, named adjusted score.

Table 5 presents the minimum, maximum, mean, and standard deviation for unadjusted evaluation scores and adjusted scores. The unadjusted scores are between 2.52 and 4.55 on a 1-5 scale. Of the 114 courses, 113 (99.1%) courses are rated above the middle (greater than 3) of the entire raw-score scale. This is phenomenon of a generosity error which leads a spurious result. The adjusted scores are between 16.45 and 74.94.

Table 4. *Summary of the Final Regression Analysis for Five Background Variables Predicting the evaluation score (N = 114)*

Variable	B	$SE\ b$	β	R	R^2	F
a=1.581				.895	.800	86.642***
Student enthusiasm	.408**	.066	.567			
Participation	.146*	.061	.224			
Expected grade	.074	.086	.083			
Grading standard	-.144*	.060	-.144			
Course difficulty	-.061	.089	-.076			

* $p < .05$ ** $p < .01$

Table 5. *The Minimum, Maximum, Mean, and Standard Deviation for Unadjusted Scores and Adjusted Scores (N = 114)*

Score	Minimum	Maximum	M	SD	N_1	N_2
Unadjusted	2.52	4.55	4.00	.302	113(99.1%)	59(51.8%)
Adjusted	16.45	74.94	50	10	57(50.0%)	57(50.0%)

Note. N_1 : the number of cases above the middle of the scales; N_2: the number of cases above the means.

Table 6 presents the number of cases below (ineffectiveness) and above (effectiveness) the means of unadjusted scores and adjusted scores. In order to make a classification for each course, the course rated lower than mean was treated as ineffective course and the course higher than mean was classified as effective course. This operational definition applies for both unadjusted scores and adjusted scores as well. Twenty unadjusted effective courses are classified as adjusted ineffective courses. On the other hand, eighteen unadjusted ineffective courses are classified as adjusted effective courses. The consistency of course- ranking classification is only 66.7% (76/114 = 66.7%). Besides, the correlation between unadjusted scores and adjusted scores is .447 and the correlation between unadjusted course rankings and adjusted course rankings is .334.

Table 6. *The Number of Cases below and above the Means of Unadjusted Scores and Adjusted Scores*

Unadjusted score	Adjusted score		Total courses
	Below mean (ineffective)	Above mean (effective)	
Below mean (ineffective)	37	18	55
Above mean (effective)	20	39	59
Total courses	57	57	114

Note. The consistency of course ranking classification is (37+39)/114 =66.7%.

There are eighteen courses which are scored as ineffective by unadjusted score and are scored as effective by adjusted score. Table 7 includes the information for these courses. Some courses scored and ranked relatively low by the unadjusted score are scored and ranked high by the adjusted scores. For example, Cases 3, 10, and 18, which were ranked as the 66th, 85th, and 111th by unadjusted score, were ranked as the 5th, 8th, and 1st, respectively.

Table 7. *The Scores and Ranks for the Eighteen Courses Which Are Scored as Ineffective by Unadjusted Score and as Effective by Adjusted Score*

Cases	Unadjusted ineffective score	Adjusted effective score	Unadjusted rank	Adjusted rank
1	3.99	54.60	60	33
2	3.98	54.31	64	35
3	3.98	67.55	66	5
4	3.97	53.23	68	37
5	3.95	57.71	70	24
6	3.94	58.38	72	19
7	3.93	62.12	74	11
8	3.90	52.38	80	42
9	3.89	62.25	82	10
10	3.86	63.87	85	8
11	3.83	56.95	88	26
12	3.81	50.35	91	56
13	3.81	54.54	92	34
14	3.80	58.32	94	20
15	3.77	50.68	98	53
16	3.66	50.19	103	57
17	3.62	50.54	105	55
18	3.40	74.94	111	1

Table 8. *The Scores and Ranks for the Twenty Courses Which Are Scored as Effective by Unadjusted Score and as Ineffective by Adjusted Score*

Cases	Unadjusted effective score	Adjusted ineffective score	Unadjusted rank	Adjusted rank
1	4.51	48.51	3	67
2	4.32	49.29	14	62
3	4.25	48.19	20	71
4	4.24	43.90	22	90
5	4.24	42.34	23	95
6	4.23	49.42	24	61
7	4.22	49.87	26	58
8	4.17	48.32	32	68
9	4.17	46.65	33	80
10	4.12	47.14	34	77
11	4.12	41.71	35	99
12	4.11	42.75	38	93
13	4.09	45.82	42	82
14	4.07	47.71	45	73
15	4.06	49.13	47	64
16	4.05	45.62	49	84
17	4.05	48.65	50	66
18	4.03	49.82	56	59
19	4.02	38.93	58	102
20	4.01	46.22	59	81

On the other hand, there are twenty courses which are scored as effective by unadjusted score and are scored as ineffective by adjusted score. Table 8 includes the information for these courses. Some courses scored and ranked relatively high by the unadjusted score are scored and ranked low by the adjusted scores. For example, Cases 1, 2, and 3, which were ranked as the 3rd, 14^{th}, and 20^{th} by unadjusted score, were ranked as the 67^{th}, 62nd, and 71^{st}, respectively.

Discussion and Conclusion

This study was to identify sources of desired or undesired influences on student ratings by using the undergraduate courses from Department of Elementary Education at Hualien Teachers College in the spring semester 1999. The linear regression model developed for the course effectiveness controlled for the "determined" variables affecting student ratings. The student rating score was regressed on all the "determined" variables. The effects of the "determined" variables were moved by subtracting the original score from the regression estimate of each score (predicted score). The residual between the unadjusted score and the predicted score was converted to T score. The T score was referred to as adjusted course effectiveness score.

The findings confirmed many of factors that earlier studies have shown to influence student ratings. Based on the results, student ratings scores can be explained about 80 percent by the five teaching unrelated variables, especially, student enthusiasm. Consistent with Prave and Bairl's (1993) study, student ratings scores can be explained most largely by student enthusiasm. However, Marsh and Dunkin (1992) suggested that student enthusiasm was better interpreted as a variable impacting the quality of education rather than a bias which is a specific to student ratings. Although the regression coefficients of expected grade and course difficulty were not statistically significant in the regression model, they were still included in the model according to the previous studies. The further study can be focused on cross-validation of the effect of these two variables on student ratings.

The results of this study indicate that what are classified as effective courses by unadjusted student ratings may not correspond with what are classified as effective courses by the adjusted score. If institutions continue to believe in the importance of student voice in evaluating faculty, it may be necessary to control for the variables, which may not, related to faculty teaching performance but inappropriately contribute to student ratings. For example, instructors teaching higher-difficulty courses with relatively strict grading standards could expect to have ratings increased by the adjustment, whereas instructors teaching lower-difficulty course giving a high proportion 90s could expect to have ratings decreased.

This study provides perspectives of course effectiveness for exploring answers to the questions related to effectively assessing course/teacher teaching performance. The course effectiveness is to refine student ratings measures by eliminating, to the extent possible, pre-existing influence or effect of factors outside the control of the faculty (such as course difficulty, student motivation). The same or similar procedure can be applied to another department or school for the determination of course effectiveness.

Student ratings systems are often distrusted and resisted by university teachers because many of them believe that students' evaluations are biased by a number of factors unrelated to course effectiveness. However, while this argument may be valid with regard to the student ratings per se, it may not hold if the concern is about the use and interpretation of student ratings for making comparative judgements, which is becoming increasingly common in higher education. The findings of this study suggest that there are some sources of potential biases when raw student ratings are used crudely for making comparative judgements of teachers across instructional contexts. As least, it is certainly not fair to the teachers and courses if they are judged by the raw student ratings they receive without taking into consideration the differences in their teaching contexts. The implication is that users of student ratings, including university teachers and administrators, should recognize their limitations and use them with extreme caution in making judgemental decision.

Continued administration of the course effectiveness would provide additional information for administrative decisions, course selection, and instruction improvement. An aggregate of multiple sections within different course effectiveness from a teacher should be applied to faculty evaluation. Longitudinal student ratings data may provide more details related to the following important questions: Are certain courses continuously ranked higher than others? Are courses taught by certain teachers continuously ranked higher than the same courses taught by others? And more importantly, what are the consequences of having implemented the system of reporting both unadjusted and adjusted rating scores?

Acknowledgments
The data in this study was provided by the Academic Administrative Office of National Hualien Teachers College.
This paper was presented at the annual meeting of the American Educational Research Association (AERA), New Orleans, LA, April, 2000. I express my appreciation to the three AERA reviewers for comments on this paper.
I express my appreciation to the two journal reviewers for comments on this paper.

References

Arreola, R. A. (1995). *Developing a comprehensive faculty evaluation system.* Bolton, MA: Anker.

Brandenburg, D. C., Slinde, J. A., & Batista, E. E. (1977). Student ratings of instruction: Validity and normative interpretations. *Research in Higher Education, 7,* 67-78.

Braskamp, L. A., & Ory, J. C. (1994). *Assessing faculty work.* San Francisco: Jossey-Bass.

Bridgeman, W. J. (1986). Student evaluation viewed as a group process factor. *Journal of Psychology, 120,* 183-190.

Brown, R. (1977). *The relationships between student evaluation of teaching, student achievement and student perception.* (ERIC Document Reproduction Service No. ED 133 314)

Centra, J. A. (1993). *Reflective faculty evaluation.* San Francisco: Jossey-Bass.

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

Chang, T. (1997). *Student ratings of instruction at teachers college in Taiwan.* Hualien, Taiwan: Graduate Institute of Compulsory Education, National Hualien Teachers College. (ERIC Document Reproduction Service No. ED 414 798)

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education, 9,* 199-242.

Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education, 18,* 3-

124.

Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses: A closer look. *Research in Higher Education, 21,* 45-116.

Feldman, K. A. (1992). College students' views of male and female college teachers: Part I –evidence from the social laboratory and experiments. *Research in Higher Education, 33,* 317-375.

Feldman, K. A. (1993). College students' views of male and female college teachers: Part II-Evidence from students' evaluation of their classroom teachers. *Research in Higher Education, 34,* 151-211.

Feldman, K. A. (1997). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Eds.), *Effective teaching in higher education: Research and practice* (pp. 368-395). Bronx, NY: Agathon.

Franklin, J. Thell, M, & Ludlow, L. (1991). *Grade inflation and student ratings: A closer look.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Greenwald, A. G., & Gillmore, G. M. (1998). How useful are student ratings? Research to comments on the current issues section. *American Psychologist, 53*(11), 1228-1229.

Jirovec, R., Ramanatha, C., & Alvarez, A. (1998). Course evaluations: What are social work students telling us about course effectiveness? *Journal of Social Work Education, 34*(2), 229-235.

Koshland, D. E. (1991). Teaching and research. *Science, 251,* 249.

Koushki, P. A., & Kuhn, A. J. (1982). How reliable are student evaluations of teachers? *Engineering Education, 72,* 362-367.

Kwan, K. (1999). How fair are student ratings in assessing the teaching performance of university teachers? *Assessment & Evaluation in Higher Education, 24*(2), 181-195.

Marsh, H. W. (1980). The influence of student, course, and instructor characteristics in evaluations of university teaching. *American Educational Research Journal, 17,* 219-237.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11,* 253-388.

Marsh, H. W., & Dunkin, M. J. (1992). Students' evaluations of university teaching: A multidimensional perspective. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (pp. 143-233). New York: Agathon Press.

Marsh, H. W., & Roche, L. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal, 30,* 217-251.

Marsh, H. W., & Roche, L. A.(1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist, 52*(11), 1187-1197.

McKeachie, W. J. (1990). Research on college teaching: The historical background. *Journal of Educational Psychology, 82,* 189-200.

Mendro, R. L., Webster, W. J., Bembry, K. L., & Orsak, T. H. (1995). *An application of hierarchical linear modeling in determining school effectiveness.* Paper presented at the meeting of the American Educational Research Association, San Francisco.

Nimmer, J. G., & 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,* 195-215.

Owie, I. (1985). Incongruence between expected and obtained grades and students' ratings of the instructor. *Journal of Instructional Psychology, 12,* 196-397.

Prave, R. S., & Baril, G. L. (1993). Instructor ratings: Controlling for bias from initial student interest. *Journal of Education for Business, 68,* 362-366.

Roche, L. A., & Marsh, H. W. (1998). Workload, grades, and students' evaluations of teaching: Clear understanding sometimes requires more patient explanations. *American Psychologist, 53*(11), 1230-1231.

Ryan, J. J., Anderson, J. A., & Birchler, A. G. (1980). Student evaluation: The faculty respond. *Research in Higher Education, 12,* 317-333.

Scherr, F. C., & Scherr, S. S. (1990). Bias in student evaluations of teacher effectiveness. *Journal of Education for Business, 65,* 365-358.

Tatro, C. N. (1995). Gender effects on student evaluations of faculty. *Journal of Research and Development in Education, 28,* 169-173.

Theall, M., & Franklin, J. (1991). Using student ratings for teaching improvement. In M. Theall & J. Franklin (Eds.), *Effective practices for improving teaching:*

New directions for teaching and learning (pp. 83-96). San Francisco: Jossey-Bass.

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

AERA

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

ERIC

TM033171

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: An Application of Regression Models with Student Ratings in Determining Course Effectiveness	
Author(s): Te-Sheng Chang	
Corporate Source: National Hualien Teachers college	Publication Date: 2001 / July

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: Te-Sheng Chang	Printed Name/Position/Title: Te-Sheng Chang / professor	
Organization/Address: National Hualien Teachers College / P.O. Box 20-76, Hualien, Taiwan 970	Telephone: 011-886-3-822-5453	FAX: 011-886-3-823-7100
	E-Mail Address: achang@sparc2.nhltc.edu.tw	Date: 2001/June/21

(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