DOCUMENT RESUME

ED 258 513 HE 018 471

AUTHOR Bruce, A. Jerry
TITLE A Comparison of Three Teaching Evaluation Instruments.
PUB DATE Apr 85
NOTE 22p.; Paper presented at the Annual Convention of the Southwestern Psychological Association (31st, Austin, TX, April 18-20, 1985).
PUB TYPE Report Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Instruction; Comparative Analysis; *Faculty Evaluation; Higher Education; *Questionnaires; *Rating Scales; Student Attitudes; *Student Evaluation of Teacher Performance; Teacher Attitudes; Teacher Effectiveness
IDENTIFIERS *College Instructional Report (Univ of Arizona); Instructional Effectiveness Assessment (Kansas); *Student Instructional Report (ETS)

ABSTRACT

Three instruments used to evaluate faculty were compared: the Student Instructional Report (SIR), produced by the Educational Testing Service, the College Instructional Evaluation Questionnaire (CIEQ), produced by the University of Arizona, and the Instructional Development and Effectiveness Assessment (IDEA), produced by Kansas State University. Information is provided on student and faculty preferences, correlations among instruments and scale scores, and content. The three instruments were administered to 426 students at 16 selected classes at Sam Houston State University. SIR appeared to measure the differential components of teaching with more clarity and was preferred by students and faculty over the other two instruments. It also had the greatest amount of feedback available. CIEQ was simpler to read, was shorter, and had fewer categories. IDEA, which seemed the most complex of the three, had many categories and much feedback, but was designed more for faculty development purposes than were the other two. In terms of cost, SIR was the most and CIEQ the least expensive. Since a high degree of correlation was found among the instruments, a single general factor underlying student ratings of instruction seemed to exist. Brief descriptions of each instrument are included. (SW)

A COMPARISON OF THREE TEACHING EVALUATION INSTRUMENTS

A. JERRY BRUCE

The evaluation of instruction/faculty by students has received much attention in the past few years. However, little attention has been paid to the comparison of the various rating instruments available. The present paper attempts to compare three of the most widely used instruments: The Student Instructional Report (ETS), The College Instructional Report (University of Arizona), and The Instructional Effectiveness Assessment (Kansas State University). Student preferences, faculty preferences, correlations among instruments and scale scores, and analyses of content are reported.

A COMPARISON OF THREE TEACHING EVALUATION INSTRUMENTS

A. JERRY BRUCE

SAM HOUSTON STATE UNIVERSITY

(Paper from the Thirty-first Annual Convention
Of the Southwestern Psychological Association,
April 20, 1985)

One of the most trying problems facing college and university administrators is that of faculty evaluation. As merit pay and other performance based approaches to faculty salaries and other reward systems are promoted, the problem of evaluation becomes even more critical. Along with increased evaluation, the role of academic evaluator becomes more visible. Most administrators are eager to have available more objective means for decision making in this relm.

Faculty perform a variety of functions within the university and college community; they do research; they perform community service; they serve on various college and university committees; and they teach. The most difficult function to measure is teaching. One university administrator was once overheard addressing his faculty on this subject saying, "Teaching is by far the most important task you perform at the university, but we cannot measure it; therefore, your promotions, salaries, etc. will be determined by something we can measure, your research and publication records." One is reminded of the story of the young boy searching for his lost coin one night under a

street light. A passerby stopped to give assistance and questioned the lad about where the coin was lost. The boy pointed into a dark alley some distance away. The surprised Samaritan asked the boy why he was looking here under the street light if in fact the coin was lost there in the alley. The young philosopher replied to the obviously less intelligent adult, "The light is better here."

Many approaches have been used in attempting to measure teaching, but by far the most widely used method is that of student evaluations (Centra, 1979). Student evaluation of faculty performance is not always popular on college campuses, but it is a reality and it is, as the research tends to show (Centra, 1979), the most reliable of the method available and, perhaps, possesses the fewest severe side effects.

In developing a program of student evaluation of faculty performance, one of the obvious problems is that of choosing an instrument. As Milton et al. (1978) points out one should not casually produce a homemade device and quickly introduce it for the purpose of making important decisions. There are many standardized instruments available so why reinvent the wheel? However, the administrator must decide which one of the many available instruments is best for his/her purpose. The present report relates some data hopefully relevant to this point.

Method

The present research compares three of the most widely used (Centra, 1979; L. M. Aleamoni, personal communication, May, 1983) instruments for student evaluation of faculty: The Student Instructional Report (SIR) produced by Educational Testing Service, The College Instructional Evaluation Questionnaire (CIEQ) produced by the University of Arizona, and The Instructional Development and Effectiveness Assessment (IDEA) produced by Kansas State University. These three instruments were administered to select classes during a summer session at Sam Houston State University.

Subjects.

There were 16 classes involved in the administration of the three instruments, a total of 426 students. The classes were selected in an attempt to represent a cross-section of the university population. The following criteria were used:

1. At least one class from the following levels:

   (a) Lower level required class of 30 to 50.

   (b) Lower level lecture class not required of 30 to 50.

   (c) Lower level lecture/lab class of 20 to 50.

   (d) Upper level lecture class of 20 to 40.

(e) Upper level non-lecture course of 10 to 30.

(f) Master's level class of 7 to 20.

(g) Doctoral level class of 5 to 10.

2. At least one course from each college within the university's organizational structure.

3. No class of less than 5.

4. Class taught by full-time regular faculty.

Instruments.

SIR. The SIR instrument is a 39 item questionnaire. In addition it contains a space for five items selected by the local faculty member. The items cover a wide variety of topics: course organization and planning, faculty/student interaction, communication, course difficulty and workload, textbooks and readings, tests and exams, overall evaluations, student and course descriptive items, local options items, and miscellaneous. Feedback is given from Educational Testing Service by percent responding to each item, item means, percentile equivalent of means, and scores (percentiles and factor scores) on six factor scores (Table 1) these factor scores are based on previously identified factors from factor analysis (Centra, 1973). Comparative data for more than 30 academic disciplines, various class sizes, school size, level of class, and type of class are available separately.

CIEQ. The CIEQ instrument contains 21 standard items plus seven openended questions on the reverse side of the form. The items cover a wide variety of topics as can be seen in Table 1. Feedback is given in the proportion and frequency of responses, means, and standard deviations on each of the 21 items. Scores are also given for five scales scores plus a total. These scales are based on previously identified factors from factor analytical research (Aleamoni, 1978). In addition there are several scores given for the scales and total comparing the individual faculty member's evaluation with other faculty evaluations from similar courses.

IDEA. The IDEA form contains 39 numbered statements and seven lettered statements. The instructor also has the option of including up to five additional items of his/her choosing. The statements are grouped into sections labeled the instructor, progress on (the student is asked to compare his/her progress in this course to other courses being taken), the course, self-rating, and the respondent characteristics. Feedback is presented to the instructor by subject matter mastery, development of general skills and personal development, i.e., how much progress the students have made in these areas. Other sections present descriptions of the course, the student's self rating, the teaching methods used, a section for additional questions, and finally a diagnostic summary. Scores are presented as frequencies of the five-point rating scale, means, a

difference score and a translation (high, high average, average, low average, low). The diagnostic summary is presented with scores on teaching methods most needing attention. The factor scores are presented in the summary profile and contain the seven categories presented in Table 1.

Procedure.

Each student completed all evaluation forms (SIR, CIEQ, IDEA, and the brief preference questionnaire) in the same class period. Each class was given a predetermined sequence of administration for the three instruments. These sequences of administration were randomly selected by the experimenter to control for order of administration.

Results and Discussion

Questionnaire Correlations.

Using the IDEA overall evaluation score, the CIEQ total score, and a composite score derived from items 38 and 39 from SIR, Spearman rank-order correlation coefficients were calculated. The results (see Table 2) were across the board extremely high. If one is using the test for evaluation and an overall score is needed, all three tests seem to do the job equally well or poorly. If one desires information for faculty development purposes then the SIR seems to this writer more appropriate.

The correlations between the various factor scores on the test were also of interest. The Spearman rank-order correlations between the seven factor scores on the IDEA ranged from .56 to .91 (see Table 3). If the exams factor score is eliminated the lowest correlation is .72. One might argue that the factor scores add little information or that good teachers are good on all accounts and vise versa. On the six factor scores of the CIEQ the same can be said (see Table 4). The lowest correlation was .81. The SIR on the other hand had factor scores that correlated at much lower levels with themselves (see Table 5), the lowest correlation being .16 and the highest .90, eliminating the composite score the range was .16 to .70. The SIR text/reading score was not reported in a majority of the cases. This score is a composite of items 32 and 33, 33 asking the students to rate the readings. Many of the classes did not have readings other than the text; therefore, the score could not be calculated. It would appear that the separate measures are measuring different elements.

Comparing the factor scores of the three instruments is somewhat difficult since each instrument has a different set of factor scores. Review of the evaluation literature by Centra (1972) revealed three common factors in most instruments: Organization or structure, teaching skills or communication, and student rapport or empathy. There was a possible fourth factor, student effort or involvement. It

is difficulty to identify these factors in the three instruments being reported here. Nevertheless, comparison of the CIEQ factor scores with factor scores of the IDEA revealed a remarkable degree of similarity (see Table 6). The correlations ranged from .75 between CIEQ method and IDEA overall to .95 between CIEQ total and IDEA creating enthusaism. The correlation between the SIR and the other two questionnaires were much more varied (see Tables 7 and 8), SIR/CIEQ ranging from .35 to .94 and SIR/IDEA from .35 to .92. It seems rather clear from these results that the SIR comes closer to differentiating characteristics than the other two.

The high degree of correlation among these instruments suggests the existence of a single general factor underlying student ratings of instruction.

Faculty/Student Preferences.

The 16 faculty and 426 students were administered a brief questionnaire. Of the 16 faculty only 10 returned their forms for a return rate of 62.5%. Of the 426 students, 332 returned the questionnaire for a return rate of 77.9%. In the questionnaire the subjects were asked which of the instruments they preferred, which they considered least difficult for the students to complete and understand, which provided the soundest judgement of faculty efforts, and finally, to the faculty only, which provided the best information for faculty evaluation and development.

From both student and faculty the SIR was rated to be superior and the IDEA was rated lowest (see Table 9). These ratings suggest a greater face validity for the SIR and that faculty and students see it as more useful.

Content Evaluation.

On all three questionnaires there seemed to be a balanced attempt to equalize the items directly related to the instructor with the items directly related to the course content. Questions on examinations were found in the SIR and IDEA but on the CIEQ the only exam question was an openended items on the back of the questionnaire. Specific questions on the textbook, readings, and laboratories were found on the SIR but only implied in the IDEA and only in the openended items of the CIEQ.

The CIEQ questions were simpler and shorter. The average number of words per questions were 7.81 for the CIEQ, 10.21 for the SIR, and 9.83 for the IDEA.

Each of the three contained items that greatly overlaped and each had items that were unique to it; however, the IDEA contained some of the more interesting unique items, e.g., "I have given thoughtful consideration to the questions on this form," "How well did the questions on this form permit you to describe your impressions of the instructor and course?" and "For how many courses have you filled out this form during the present term?"

One other consideration of content regarded the question of student acquiescent tendencies, a well known consideration in questionnaire construction. A good questionnaire should have a balance of yes/no responses so that yes and no both will be used to indicate a positive evaluation of the person or concept under consideration. The SIR, in order to indicate that the instructor did a good job, only in two cases was the respondent required to indicate no; on the IDEA four times; but on the CIEQ 11 times. The CIEQ was the only instrument that seriously attempted to balance the yes/no reponses.

Conclusion

It is difficult to say which of these questionnaires would be best for a specific situation without examining carefully the needs of that specific situation. Nevertheless, one might say that the SIR appeared to measure the differential components of teaching with more clarity, the SIR was preferred by students and faculty over the other two instruments, and the SIR had the greatest amount of feedback available. On the other hand the CIEQ was simpler to read, it was shorter, and it had fewer categories. The IDEA seemed the most complex of the three. IDEA had a multitude of categories of items and feedback, but it was design more for the purpose of faculty development perhaps than the other two.

One last point, the issue of cost, on this the SIR won "going away." The SIR was much more expensive than the other two instruments. The CIEQ was the cheapest.

The evaluation of teaching is not an easy task. As Tucker (1984) says, "The art of evaluating the performance of faculty members is not that well developed" (p. 151). But it is an important task and one that needs increasing effort from the research community. This report is meager and filled with many problems but it is a attempt to begin the evaluation of the available instruments.

References

Aleamoni, L.M. (1978). Development and factorial validation of the Arizona Course/Instructor Evaluation Questionnaire. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 38, 1063-1067.

Centra, J.A. (1972). The Student Instructional Report: It's development and use. (SIR Report # 1). Princeton, NJ: Educational Testing Service.

Centra, J.A. (1973). The Student Instructional Report: Comparison with alumni ratings, item reliability, and the factor structure. (SIR Report # 3). Princeton, NJ: Educational Testing Service.

Centra, J.A. (1979). DETERMINING FACULTY EFFECTIVENESS. San Francisco: Josey-Bass.

Milton, O. & Associates. (1978). ON COLLEGE TEACHING. San Francisco: Josey-Bass.

Tucker, A. (1984). CHAIRING THE ACADEMIC DEPARTMENT (2nd Ed). New York: American Council on Education/Macmillian.

TABLE 1

COMPARISON OF EVALUATION INSTRUMENTS' SUBSCORES

SIR	Faculty/student interaction
	Communications
	Tests, exams, textbook, and readings
	Course organization and planning
	Course difficulty and workload
	Student interest
	Overall
CIEQ	General attitude
	Method (of instruction)
	Content
	Interest
	Instructor
	Total
IDEA	Outcomes
	Overall evaluation
	Would like instructor again
	Improved attitude toward field
	Method
	Involving students
	Communicating content and purpose
	Creating enthusiasm
	Preparing examinations

TABLE 2

COMPARISON OF OVERALL EVALUATIONS ON CIEQ, IDEA, SIR

	CIEQ	IDEA	SIR
CIEQ	--		
IDEA	.89	--	
SIR	.94	.88	--

TABLE 3

IDEA SUBSCORE CORRELATIONS

OUTCOME						
overall	--					
would like instructor again	.72	--				
improved attitude toward field	.74	.91	--			
METHOD						
involving students	.86	.81	.64	--		
communication of content & purpose	.86	.80	.77	.83	--	
creating enthusaism	.86	.90	.78	.92	.81	--
preparing exams	.70	.68	.56	.82	.71	.70	--

TABLE 4

CIEQ SUBSCORE CORRELATIONS

General attitude	--					
Method	.81	--				
Content	.81	.87	--			
Interest	.94	.82	.85	--		
Instructor	.82	.88	.83	.84	--	
Total	.92	.93	.88	.92	.95	--

TABLE 5

SIR SUBSCORE CORRELATIONS

Course organization & planning	--						
Faculty/student interaction	.58	--					
Communication	.29	.70	--				
Course difficulty & workload	.33	.58	.39	--			
Textbook & readings	--	--	--	--	--		
Test & exams	.65	.59	.26	.16	--	--	
Overall score	.54	.90	.83	.43	--	.69	--

TABLE 6

CIEQ/IDEA SUBSCORE COMPARISONS

	general attitude	method	content	interest	instructor	total
OUTCOME						
overall	.88	.75	.92	.90	.82	.89
would like instructor again	.81	.98	.81	.84	.91	.93
improved attitude toward field	.86	.89	.80	.82	.81	.86
METHOD						
involving students	.79	.82	.85	.87	.87	.91
communication of content & purpose	.76	.84	.90	.78	.80	.85
creating enthusaism	.87	.88	.84	.94	.90	.95
preparing exams	.69	.75	.77	.63	.80	.80

TABLE 7

IDEA/SIR SUBSCORE CORRELATIONS

	course organization & planning	faculty/ student interaction	communication	course difficulty & workload	text & readings	tests & exams	overall score
OUTCOME							
overall	.37	.73	.41	.43	--	.60	.88
would like instructor again	.75	.77	.67	.52	--	.57	.82
improved attitude toward field	.63	.57	.43	.52	--	.52	.70
METHOD							
involving students	.45	.91	.77	.47	--	.57	.95
communication of content & purpose	.55	.75	.50	.57	--	.62	.79
creating enthusaism	.57	.79	.65	.35	--	.55	.92
preparing exams	.51	.92	.60	.48	--	.66	.89

TABLE 8

CIEQ/SIR SUBSCORE CORRELATIONS

	general attitude	method content		interest instructor		total
course organization & planning	.42	.69	.52	.47	.71	.59
faculty/student interaction	.63	.81	.80	.68	.88	.85
communication	.51	.66	.53	.59	.59	.62
course difficulty & workload	.40	.58	.46	.35	.53	.56
textbook & readings	--	--	--	--	--	--
tests & exams	.53	.61	.75	.59	.56	.57
overall score	.86	.84	.87	.89	.90	.94

TABLE 9

FACULTY/STUDENT OPINION QUESTIONNAIRE RESULTS

Questions			CIEQ	IDEA	SIR	No Difference
prefer	faculty	n	3	0	6	1
		%	(30.0)	(0.0)	(60.0)	(10.0)
	students	n	76	37	189	30
		%	(22.9)	(11.2)	(57.0)	(9.1)
least difficult for student to:						
complete	faculty	n	1	1	7	1
		%	(10.0)	(10.0)	(70.0)	(10.0)
	students	n	71	16	196	49
		%	(21.4)	(4.9)	(59.1)	(14.8)
understand	faculty	n	1	0	6	3
		%	(10.0)	(0.0)	(60.0)	(30.0)
	students	n	66	16	191	59
		%	(19.9)	(4.9)	(57.6)	(17.8)
allowed the student to provide soundest judgement	faculty	n	3	0	6	1
		%	(30.0)	(0.0)	(60.0)	(10.0)
	students	n	95	41	154	42
		%	(28.7)	(12.4)	(46.4)	(12.7)
provides information necessary for faculty:						
evaluation	faculty	n	3	0	5	2
		%	(30.0)	(0.0)	(50.0)	(20.0)
development	faculty	n	2	1	4	3
		%	(20.0)	(10.0)	(40.0)	(30.0)