

Measuring Student's Perceptions of Teaching: Dimensions of Evaluation

Charles A. Burdsal and John W. Bardo

Results from application of the SPTE Questionnaire from 2,115 classes were factor analyzed to oblique simple structure. A first-order solution with six factors resulted: 1) Attitude Toward Students, (2) Work Load, (3) Course Value to Students, (4) Course Organization/Structure, (5) Grading Quality, and (6) Level of Material. Two second-order factors were also found. At the second-order, results showed that course difficulty was uncorrelated with other aspects of students' perceptions of teaching. Results are discussed in relation to broad concepts of teaching evaluation as well as methodological requirements of adequate assessment of teaching quality.

Assessment of the structure, reliability and validity of student evaluations of teaching performance is the subject of a vast and increasing literature in educational measurement. Typically, assessments involve students completing questionnaires consisting of one or more fixed response format "scales." These scales are then often analyzed in relation to some objective criterion of instructional quality. Also, a great deal of attention has been paid to relationships between questionnaire content and abstract models of the construct "quality teaching." One of the major issues in this literature involves the underlying dimensions of students' perceptions of teaching as well as any interrelationships among these dimensions. In this paper, major issues concerning the measurement of facets of student perceptions of teaching quality were examined.

There is general agreement in the theoretical literature that quality teaching is multifaceted: there is no unitary goal for teaching and many phenomena contribute to the perceived quality of a class or particular instructor (McKeachie, 1979). In their analysis of 21 studies of teaching effectiveness, Wotruba and Wright (1975) found ten general criteria. These include communication skills, attitudes toward students, knowledge of the subject, organization of subject matter, enthusiasm for the subject, fairness in grading, flexibility, encouragement of students to think for themselves, and ability to speak.

There have been a variety of approaches used in the empirical literature to assess the quality of teaching. Unfortunately, many of them suffer from methodological or conceptual shortcomings which lead to questions concerning the efficacy of current measurement practices. For example, while most theorists agree that teaching is a multidimensional phenomenon, many studies operationalize quality of teaching in terms of a single item, a summated scale, or a series of items which are analyzed separately. For instance, Peterson and Cooper (1980) used a single item "scale" to assess overall course quality as well as five individual item scales to examine student response to various aspects of teaching. (See also Meredith, 1982.) In a much more problematic analysis, Vasta and Sarmiento (1979) used 50 individual "evaluative items" to examine relationships between changing grading criteria and perceptions of faculty performance. They used modified grading practices in two sections of a class; one was graded harshly and the other more standardly. In total, seven of 50 items showed differences significant at $p < .05$ between the two treatments roughly three of which could be expected by chance. No multivariate test of significance was used to determine whether the overall pattern of differences in response between the two groups obtained multivariate significance. No controls in the experiment were made for obvious confounding factors such as interaction among students in the two treatment groups or previous student experience with "standard" university grading practices.

The use of individual items to assess teaching efficacy suffers from a basic methodological flaw which has long been recognized in psychometric testing: it is difficult to determine the reliability of items used individually, and, single items tend to be less correlated with their criteria than are multiple item

tests with acceptable homogeneity (see footnote in Lester, 1982; Nunnally, 1978). In other words, the relationship between a respondent's "true score" and measured "observed score" is lower for individual items than for summated tests. Based on generalized theory, the validity of individual item measurements is *a priori* more suspect than those measurements based on multiple items (Kane, Gillmore, and Crooks, 1976; Smith, 1979).

With summated tests there is another set of problems. Traditionally in measurement theory, an attitude is defined as relating to a construct with a particular domain of content (Nunnally, 1978). When the domain is complex and composed of divergent subconstructs, as is student perceptions of teaching quality, interpretation of summated test scores can become problematic. Individual or group means provide information about overall levels of affective components of attitudes, but they suggest nothing about the structure of variance within the test which produced the means (see Bardo and Foddy, 1980). In measuring a complex phenomena such as teaching quality this limitation becomes crucial. For instance, it has been reported that relationships between expected grade and teacher evaluation are unclear (Abrami, Dickens, Perry, and Leventhal, 1980; Brown, 1976; Feldman, 1976; Marsh, 1980). Part of this lack of clarity may result from the tendency of researchers to summate across subscales and not to examine relationships among subcomponents of teaching. That is, since not all scales used to evaluate teaching measure the same subconstructs or weight the subcomponents in a similar way, part of the disparity may be lack of comparability in measurement procedures.

Recently, and partly in response to the kinds of issues raised above, there has been a movement to multivariate procedures for measuring instructional quality. Various forms of factor analysis have become particularly popular. While this represents a considerable improvement in procedure, there are still major problems with the literature in choice of algorithm and reporting of results. In reviewing recent factor analytic studies of student evaluations of teaching, several major problems became apparent. For instance, some studies do not report the method of factor extraction employed or they cite previous and sometimes even unpublished literature (Marques, Lane and Dorfman, 1979; Comer, 1980). In these cases it is difficult to interpret the adequacy of solutions on which the studies are based. In other instances, subscales are built on *a priori* results from other populations (Orpen, 1980). However, the most common flaws in the factor literature involve forcing items into uncorrelated (orthogonal) solutions and the use of inadequately documented or reported oblique factor solutions.

Numerous studies have used varimax orthogonally rotated factor matrices as the basis for "determining dimensions" of teaching evaluation (e.g., Meier and Feldhusen, 1979; Bennett, 1982; Whitely and Doyle, 1979; DeCette and Kenney, 1982). In these cases, a solution in which resulting subscales are uncorrelated with each other is forced on the data. In situations where scales can be assumed on theoretical grounds to be uncorrelated, this approach may be justified. Another problem with orthogonal rotations is that, while the factors will be uncorrelated, the resulting scales or factor scores usually are not (Cattell, 1978). In the case of teaching evaluation, this assumption of orthogonality among subscales seems particularly inappropriate and contrary to existing theoretical literature (McKeachie, 1979).

Various oblique factor solutions represent an improvement in procedure, but here again there are major problems. Although there is not total agreement on the most appropriate form of oblique factor analysis and there are continuing debates over the most efficacious specific technical procedure to be used, it is widely accepted among specialists that: (1) factor analysis should be directed toward achievement of "simple structure" (Cattell, 1978; Gorsuch, 1974; Marradi, 1981; Thurstone, 1947) and (2) regardless of the specific technical solution chosen, factor correlations should be reported as a means of assisting in interpretation of results. In many studies using oblique solutions neither of these conditions are met. For instance, Bryant, Comisky, Crane, and Zillman (1980) reported that they performed "... a principal components analysis followed by factor rotation". Marsh and Overall (1980) used a "principal components solution followed by an oblique rotation." In other studies, Marsh, Overall, and Kessler (1979)

and Marsh (1982) employed "direct oblimin" solutions. In the Marsh and Overall and Bryant, et al., papers, the method of factor analysis is vague and not supported by clear descriptions of procedures.² In the Marsh (1982) and Marsh, Overall, and Kessler (1979) papers, the procedure used is sufficiently clearly presented, but adequacy of the solution still cannot be determined due to the failure to report factor correlations and significance of simple structure. While these may seem to be somewhat esoteric technical issues their effect on adequately interpreting empirical results is devastating. For instance, if the correlations of factors in an oblique solution are not reported, there is no way of determining if certain factors may be orthogonal to one another (which in the case of teaching evaluation may be of manifest importance). It also cannot be determined if the factor solution "went far enough" in obtaining a reasonably clean empirical representation of the data unless simple structure is examined.

Because of the methodological limitations of many of the factorial studies, there is a need for further development of empirically-based analyses of the domain and structure of students' perceptions of teaching if such perceptions are to play any role in evaluating "good" teaching. In the remainder of this paper, the results of a first- and second-order factor analytic study of students' perceptions of teaching performance will be presented.

Method

Instrument

In 1975 the Liberal Arts and Sciences Teaching Improvement Committee (LASTIC) was founded with the purpose of increasing the instructional quality at Wichita State University. As part of their assignment, a questionnaire was developed with the purpose of assessing student perceptions of teaching effectiveness (SPTE). The instrument since that time has undergone only minor revision (the addition of three items).

The SPTE questionnaire consists of 11 demographic items and 39 questions aimed at evaluating various aspects of teaching performance of faculty. All items are in Likert type format. In addition the students are allowed to provide open ended comments.

Subjects

The questionnaire was made available on a voluntary, confidential basis to all faculty; however, the primary use was in the College of Liberal Arts and Sciences. The results of the questionnaire were intended to be used only for feedback to the individual faculty member. Each faculty member could, at his/her discretion use the results for tenure, promotion, or salary consideration.

During the period from fall, 1977, through the spring of 1982, a total of 42,019 SPTE questionnaires were collected on 2,115 sections. Of the 2,115 sections, 66 were from the College of Business, 18 from the College of Education, 74 from the College of Engineering, 18 from the College of Fine Arts, 17 from the College of Health Related Professions, 1,351 from the College of Liberal Arts and Sciences, and 571 were unidentified.

² In the Marsh and Overall paper, SPSS oblique rotations are used. This procedure, while widely available, is problematic. Inadequate documentation is provided as to what procedures are actually used; simple structure (in terms of hyperplane count) is not presented; the program does not automatically report interfactor correlations; and no description is given as to which matrix is output for interpretation (e.g., the factor pattern or factor structure matrix). As a result, it is not possible to determine the adequacy of results reported.

Procedure

Toward the end of each semester, each faculty member choosing to use SPTE requested the appropriate number of questionnaires from the testing center. Printed instructions told the instructor after assigning a student as a proctor to administer SPTE, to leave the room. After the students filled out the forms, they were carried by the student administering the questionnaire to a central location on campus for analysis.

Analysis

As the purpose of SPTE was to measure the students' perceptions of an instructor's performance in a section, the unit of measurement chosen was the section. Item means were computed for each section eliminating missing responses from each mean. Next, an oblique factor analysis was performed upon the data from the 2,115 sections as described in the Results section. Finally, a second order factor analysis was performed upon the correlations found among the factors in the first factor analysis.

Results and Discussion

Correlation coefficients were computed for all pairwise combinations of the 39 evaluative items. Bartlett's chi-square for the significance of the variance in the correlation matrix was 95703.7 with 741 degrees of freedom yielding an equivalent z of 234.2 which was highly significant ($p < .00001$). All eigenvalues were then extracted from the resulting correlation matrix. Cattell's (1978) scree test indicated six factors which were then extracted using an iterative principle axis solution until the communalities stabilized in the third decimal place. The factor matrix was rotated by a VARIMAX orthogonal rotation followed by a PROMAX oblique rotation, three graphical hand rotations and a MAXPLANE cleanup oblique rotation yielding a 68.8% .10 hyperplane count in the reference vector structure (Vrs). All factors achieved significant simple structure at the .0001 level or better (Cattell, 1978). The resulting factor pattern (Vfp) may be found in Table 1.

Two factors as indicated by the scree test (Cattell, 1978) were then extracted from the factor correlation matrix utilizing the same procedures as described for the first order factoring except that with two factors, no graphical rotations were required. The second order factor pattern and factor correlation matrix may be found in Table 2.

Finally, to ascertain an estimate of the reliability of the factors (and for other research purposes), factor scores were calculated for each factor (both first and second order) using the regression method. Also the correlations of the factor scores with the factors were computed. The correlations of a set of factor scores with their associated factors give an estimate of common variance and thus an estimate of the reliability of the factor. The matrix of such correlations for the first and second order factors may be found in Table 3.

In examining the results, it should be emphasized that this research was focused on the dimensions of student perceptions of teaching. It is important to differentiate between "good" teaching and student's perceptions of teaching. In order to examine the relationships between such student perceptions and other teaching criteria, two prior conditions must be met: (1) it should be recognized that student perceptions of teaching quality and other academic models of "good" teaching are not necessarily identical and (2) adequate measures of student perceptions must be developed before the relationship between any two conceptions can be adequately evaluated. It appears that the results of this factor study are a step toward this end. Using a very large sample and sophisticated, technically appropriate factor analytic procedures, a solution with highly reliable factors (none below .85 as shown in Table 3) was obtained. Furthermore, the factors make good conceptual sense (a summary of each factor's salient loadings [greater than 1,351] may be found in Table 4.) Each first order factor was interpretable in terms

of theoretically significant dimensions of teaching quality. In essence, items loading each factor were consistent and logical in direction of loading.

Particularly significant are the relationships among factors at the second order. The second order factor analysis revealed two essentially orthogonal factors: one of general quality of instruction and one of difficulty.

The primaries which constitute the second order general quality factor in descending order of their weights were: (1) course organization/structure, (2) course value to the student, (3) attitude toward students, and (4) grading quality.

TABLE I
Factor Pattern (*V_{fp}*) for the Primary Factors and the Factor Correlations (*R_f*)

Item	Factor					
	I	II	III	IV	V	VI
1	.27	.11	-.01	-1.00	-.18	.02
2	.06	.69	-.10	.04	.11	.70
3	-.08	.11	-.14	-.74	-.04	-.05
4	.17	.09	-.11	-.70	-.34	-.07
5	-.13	-.11	.47	.45	.12	-.09
6	.14	.02	-.60	-.13	-.08	.09
7	.02	-.10	-.18	.38	.55	-.10
8	-.04	-.90	-.03	.15	-.02	.07
9	-.04	-.23	-.46	-.01	.00	.38
10	-.16	.03	-.06	.10	.10	-.75
11	.05	-.12	-.10	-.17	-.71	-.03
12	-.17	-.12	-.36	-.18	-.11	-.11
13	.02	.10	.13	-1.02	-.04	.08
14	-.87	.01	.13	.01	-.20	.08
15	.18	-.06	.18	.49	-.07	-.17
16	-.34	-.12	-.33	-.32	.14	-.06
17	-.52	.03	-.16	-.32	.09	.06
18	.61	.23	.05	.02	.28	.12
19	-.47	.03	.01	-.53	.07	.00
20	.38	-.02	-.02	.60	.01	.03
21	-.80	-.08	.10	.04	-.11	-.05
22	.17	-.02	.09	.53	.28	.05
23	.48	.17	-.01	.00	.31	.03
24	.04	.08	-.08	-.39	-.52	-.06
25	-.05	.19	-1.09	.18	.10	-.02
26	-.09	-.10	.85	.06	.07	-.18
27	-.07	-.52	-.11	.09	.13	.40
28	.18	-.54	.18	-.37	.39	-.08
29	-.24	.12	-.72	-.06	.10	-.54
30	.89	.04	.02	-.04	.08	-.01
31	-.03	-.08	-.80	-.05	-.03	-.25
32	-.09	.09	-.57	.12	-.35	-.05

33	-.04	-.28	.05	.00	-.13	.51
34	-.69	.04	-.07	-.17	-.10	.04
35	-.01	.00	.88	-.01	.02	.06
36	-.06	.21	-.75	-.14	-.03	-.05
37	.18	-.07	.06	.77	.00	.06
38	-.42	.11	-.10	-.46	-.06	-.01
39	-.36	.03	-.09	-.59	.00	-.01
	I					
I	1.00	II				
II	-.11	1.00	III			
III	.72	.17	1.00	IV		
IV	.75	.18	.78	1.00	V	
V	.60	-.06	.61	.57	1.00	VI
VI	.04	-.52	-.04	-.15	.13	1.00

TABLE 2
Second Order Factor Pattern (Vfp) and Factor Correlations (Rf)

	1	2
	1	2
I	-.85	-.14
II	-.06	.81
III	-.88	.10
IV	-.89	.18
V	-.69	-.15
VI	.01	.65

$r_{12} = .005.$

The course organization/structure factor was composed of those items related to the general mechanics of the classroom and the instructor's ability to handle those mechanics. Here such aspects as presentation, organization, explicitness of objectives, and ability to handle questions dominated. It is interesting to note that one item often used as a single item indicator (the ranking of the instructor) loaded on this factor pointing out that high general ranking of instructors was related to their classroom organization and structure.

On the other hand, the ranking of the course in terms of recommending it to another student was related to the perceived value of the course to the student (the second primary constituting the general quality factor). Degree of stimulated interest, judgment as to the usefulness, and increased knowledge in the subject all defined the course value factor. Here, the ability of the instructor to interest the student in the area of study appeared to be related to the student's general excitement about the subject matter of the course. If the student's interest was stimulated, he/she saw the course as useful, retained the material, and enjoyed going to the class.

TABLE 3
Correlations of Factor Scores with Factors

Primary Factors						
	I	II	III	IV	V	VI
I	.98	-.10	.74	.76	.64	.05
II	-.10	.95	.16	.17	-.06	-.58
III	.74	.17	.98	.79	.65	-.03
IV	.77	.18	.80	.99	.62	-.16
V	.61	-.06	.62	.59	.94	.14
VI	.04	-.55	-.03	-.15	.14	.90
Second Order Factors						
			1	2		
		1	.96	.00		
		2	.00	.86		

TABLE 4
Summary of the Factors by Salient Loadings

<i>Primary Factors</i>	
Factor I: Attitude Toward Students	
.89	From my own experience, the instructor came across as a person as well as a teacher (very well)
-.87	Based upon your own experience the instructors attitude toward students was (respectful)
-.80	With respect to students' freedom to express opinions and ask questions in the classroom, the instructor (freely permitted comments)
-.69	I like the instructor (very much)
.61	With respect to your progress in this course. the instructor was (concerned and actively helpful)
-.52	As reflected in the classroom and in the presentation of course material, the instructor was (very enthusiastic)
.48	For the purpose of explanation and clarification outside of the formal class period, the instructor was (readily available)
-.47	In terms of voice level, rate of speaking, appearance, and mannerisms, the personal characteristics of the instructor were judged to be (conducive to learning)
-.42	On the basis of performance in this class, would you recommend this instructor to another student? (very highly)
.38	The ability of the instructor in handling questions and answering them to the student's satisfaction was (quite satisfactory)
-.36	Considering all the instructors I have had. on the basis of general quality of instruction, I would rank this instructor as (among the best)
Factor II: Work Load	
-.90	Considering other courses of similar credit and level, the work load for this course was (very heavy)
.69	As an aid to learning, the number and difficulty of assignments were (excessive)
-.54	My feeling is that the instructor graded (very hard)

-52 With respect to my ability and prior preparation, the level of difficulty of this course was (very high)
Factor III: Course Value to Students
-1.09 As a result of this course my interest in pursuing additional knowledge in this area has been (stimulated)
.88 In one way or another (whether in relationship to my major, other courses, or just life in general) this course has been (very useful and worthwhile)
.85 As a result of this course my knowledge level in this area has (greatly increased)
-.80 Compared with other courses I have taken, I feel my retention of material, concepts, applications, etc., in this course will be (high)
-.75 On the basis of the subject material, would you recommend this course to another student (highly)
-.72 I have usually gone to the meetings of this class with (eager anticipation)
-.60 The learning resources, including the text and all other required sources of content beyond the classroom presentation were (of great value)
-.57 In terms of what I have gained from the course. the grade I expect to obtain is (a good reflection)
.47 The degree to which the material covered in this course was interrelated and consistent with subject area was (very high)
-.46 Considering the level of the course, class composition, prerequisites, etc., the level of the material presented was (very high)
-.36 The instructor's use of, or directions of students to, nonrequired references or resources was (appropriate)
Factor IV: Course Organization/Structure
-1.02 The instructor's classroom presentation was (well prepared at all times)
-1.00 Considering the nature of the subject, the entire course (was extremely well organized)
.77 In conveying the concepts of this course in a clear. meaningful, and appropriate way, the instructor's ability was (very evident)
-.74 Considering the nature of the course in terms of subject and class size, the method of presentation of material, i.e., lecture. discussion groups, etc., was (most appropriate)
-.70 The general objectives of the course were (clearly understood)
.60 The ability of the instructor in handling questions and answering them to the student's satisfaction was (quite satisfactory)
-.59 Considering all the instructors I have had, on the basis of general quality of instruction, I would rank this instructor as (among the best)
-.53 In terms of voice level, rate of speaking, appearance, and mannerisms, the personal characteristics of the instructor were judged to be (conducive to learning)
.53 The instructor helped the class achieve the objectives set forth in the course (very satisfactorily)
.49 Judging only on the basis of your own experience, the instructor's knowledge of the subject matter of the course appeared to be (exceptional)
-.46 On the basis of performance in this class, would you recommend this instructor to another student (very highly)
.45 The degree to which the material covered in this course was interrelated and consistent with subject area was (very high)
-.39 The examination questions, or other evaluative methods used by this instructor, seemed (very clear and fair)
.38 The method of assigning grades was (clearly understood and consistent)

-.37 My feeling is that the instructor graded (very hard)
-.71 The number and type of evaluations, i.e., exams, assignments, papers, etc., used in determining the grade were (sufficient to reflect achievement)
.55 The method of assigning grades was (clearly understood and consistent)
-.52 The examination questions, or other evaluative methods used by this instructor, seemed (very clear and fair)
.39 My feeling is that the instructor graded (very easy)
-.35 In terms of what I have gained from the course, the grade I expect to obtain is (a good reflection)
Factor VI: Level of Material
-. 75 Considering the nature of the course and subject material, the rate of coverage of the material was (too slow)
.51 With respect to the goals of the course the amount of material presented was (insufficient)
.40 With respect to my ability and prior preparation. the level of difficulty of this course was (very easy)
.38 Considering the level of the course, class composition, prerequisites, etc., the level of the material presented was (very low)
<i>Second Order Factors</i>
Factor I: General Quality
-.89 Course Organization/Structure (poor)
-.88 Course Value to Student (low)
-.69 Grading Quality (poor)
Factor 2: General Difficulty
.81 Work Load (high)
-.65 Level of Material (hard)

Note: For psychometric purposes, the socially desirable end of the Likert scale varied for each item. The parenthetical statement accompanying each item matches the direction indicated by the sign of the factor weight for that item.

The instructor's attitude toward students was the next primary factor, in order of weight, defining the general quality second order factor. It is composed primarily of items relating to the instructor personally. How the instructor treated the student, how open instructor was in class, his/her involvement with the subject matter, and availability all seemed important to the student's perceptions of the quality of the class. Finally, the quality of grading was related to general quality. Adequate number and type of evaluations, explicitness of grading procedures, and the perceived fairness of such procedures all were significant considerations to the student.

The second higher order factor of general difficulty was composed of two primary factors: work load and level of material. It is interesting to note that quality as rated by students was unrelated to their perceptions of course difficulty ($r = .005$).

Conclusions

The overall factor structure lends several comments concerning some current practices in measuring student perceptions of teaching. First, the idea of using single item indicators is further challenged. Three such items were included in the SPTE form (overall course evaluation, general ranking of the instructor, and the degree to which a student would recommend the instructor to other students). While the items had significant mutual intercorrelation, all had their highest loadings on different primary factors. This gives strong evidence that such single items are overly simplistic representations of students' perceptions of the instructor. This problem is exacerbated if these items are viewed as being unidimensional measures of teaching quality, since the reliability of single items (in terms of response

homogeneity) is undefinable and because the measurement literature, whether based upon reliability or generalizability theory, stresses the inadequacy of single items as indicators of a complex theoretical referent. Thus, using single, omnibus items is a dubious practice and should be avoided. In terms of administrative decisions, the use of single item indicators is risky and questionable especially when used for tenure, promotion, or salary considerations.

The desire for a single "score" by which teaching can be "evaluated" is reminiscent of the misuse of IQ scores for school placement, consideration for special education, etc. It is somewhat amazing that with all the criticism of such scores, which are in most cases of good psychometric quality, psychometrically poor single item indicators are being used as at least a major part of evaluating teaching. What would be the reaction to evaluating IQ with a single item from the WAIS or for that matter basing a student's semester grade on a single item examination? The idea of placing a child in a special education program when his IQ is 79 but not when his IQ is 81 is parallel to and probably more accurate than evaluating an instructor as being a good teacher if he/she averaged 4.5 on one of the single item indicators but classifying the individual as a poor teacher if the mean on the item were 4.1.

Second, it should be noted that the results highlight another problem in the teaching evaluation literature: the use of nonfactored summated scales. It is generally recognized that even tests (or scales) with high reliability, as measured by internal consistency type reliability coefficients such as Cronbach's alpha, may be factored into several subdimensions. As a result, summated tests tend to combine in one measure variance that should be partitioned into several more specific indicators if the primary purpose of the measurement is to provide insight into the domain being studied. This may partly explain contradictory findings of relationships between students' perceptions of teaching quality and other variables such as expected grade. In this study, course difficulty was found to be unrelated to the second order dimension of teaching quality. If expected grade were correlated with course difficulty but not with overall quality, and if one produced a summated scale of "teaching quality" which included items from both the overall quality factor and the course difficulty factor, one would find a correlation between expected grade and "teaching quality" resulting in inappropriate and erroneous conclusions. Extraneous variables relating only to course difficulty could be spuriously interpreted to predict overall teaching quality. Again, the results of this study highlight the need for greater attention to construction and analysis of all measures of teaching quality.

Third, part of the problem in evaluating teaching comes from the difficulty in defining "good teaching." University faculty, who are most involved in debating teaching evaluation, often find such evaluation threatening especially considering its potential use for tenure, promotion, and salary decisions. The nebulous conceptualization of "good teaching," both in the theoretical and empirical literature, allows objections to develop (many valid) for whatever approach is taken to such evaluation (Cahn, 1982). As a consequence, it becomes easy to argue that any teaching approach is "good" and generally acceptable. However, there is an increasing pressure from external sources, especially those providing financial support to universities and colleges, for such evaluation. Single item indicators whose face validity is high seem to answer many of the faculty objections to evaluation. For example, many faculty believe that students are incapable, in any case, of evaluating teaching and at best only a general impression from students provide any useful information (Cahn, 1982).

Such a procedure is risky at best as shown by the criticisms presented in the introduction and the fact that such items load different factors. A better approach is to take students' perceptions of teaching for what they are: students' perceptions of teaching. It is generally agreed that "good" teaching is multifaceted. It thus seems logical to measure students' perceptions of teaching as accurately as possible using good psychometric practices and then treating such measurements as one aspect of "good" teaching. After all, it is the student who has to live with our teaching. As students' views of teaching are not unidimensional, this avoids the problem (to some the advantage) of a single indicator of "good" teaching. It also provides a fuller view of an instructor's performance in the classroom. In essence, the

various dimensions of students' perceptions of teaching need to be identified, accurately measured, and examined in relationship to other criteria of "good" teaching.

In sum, the logical patterns in which the items from SPTE fall give face validity to student evaluation, the fact that students' perceptions of what they consider quality is unrelated to the course difficulty is encouraging. How easy the instructor graded had only a minor loading (.39) on the quality of grading factor which had the lowest salient loading on the general quality second order factor. The results support a multidimensional perspective of students' perceptions of teaching; also emphasized is the need for theoretically justified and methodologically sound analyses of teaching performance.

REFERENCES

- Abrami, P. C., Dickens, W. J., Perry, R. P., and Leventhal, L. (1980). Do teacher standards for assigning grades affect student evaluations of instruction? *Journal of Educational Psychology, 72*, 107-118.
- Bardo, J. W. and Foody, W. H. (1980). On interpreting internal consistency-type reliability in Likert-type summated rating scales. *Sociological Forum, 3*, 75-95.
- Bennett, S. K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching. *Journal of Educational Psychology, 74*, 170-179.
- Brown, D. L. (1976). Faculty ratings and student grades: A university-wide multiple regression analysis. *Journal of Educational Psychology, 68*, 573-578.
- Bryant, J., Comisky, P. W., Crane, J. S., and Zillman, D. (1980). Relationship between college teachers' use of humor in the classroom and students' evaluations of their teachers. *Journal of Educational Psychology, 72*, 511-519.
- Cahn, S. M. (1982). The authority of the teacher. *Academe*, May-June, 13-14.
- Cattell, R. B. (1978). *The scientific use of factor analysis in the behavioral and life sciences*. New York: Plenum Books.
- Comer, J. C. (1980). The influence of mood on student evaluations of teaching. *Journal of Educational Research, 73*, 229-232.
- DeCette, J. and Kenney J. (1982). Do grading standards affect student evaluations of teaching? Some new evidence on an old question. *Journal of Educational Psychology, 74*, 308-314.
- Feldman, K. A. (1976). Grades and college students' evaluations of their courses and teachers. *Research in Higher Education, 4*, 69-111.
- Gorsuch, R. (1974). *Factor analysis*. Philadelphia: Saunders.
- Kane, M. T., Gillmore, G. M., and Crooks, T. J. (1976). Student evaluations of teaching: The generalizability of class means. *Journal of Educational Measurement, 13*, 171-184.
- Lester, D. (1982). Students' evaluation of teaching and course performance. *Psychological Reports, 50*, 1126.
- Marques, T. E., Lane, D. M., and Dorfman, P. W. (1979). Toward the development of a system for instructional evaluation: Is there consensus regarding what constitutes effective teaching? *Journal of Educational Psychology, 71*, 840-849.
- Marradi, A. (1981). Factor analysis as an aid in the formation and refinement of empirically useful concepts. In D. J. Jackson and E. F. Borgatta (eds.). *Factor analysis and measurement in sociological research: a multi-dimensional perspective*. Beverly Hills: Sage, 11-49.
- Marsh, H. W. (1982). Validity of students' evaluations of college teaching: A multitrait-multimethod analysis. *Journal of Educational Psychology, 74*, 264-279.
- Marsh, H. W. and Overall, J. U. (1980). Validity of students' evaluations of teaching effectiveness: Cognitive and affective criteria. *Journal of Educational Psychology, 72*, 468-475.

- Marsh, H. W., Overall, J. U. and Kesler, S. P. (1979). Validity of student evaluations of instructional quality: A comparison of faculty self-evaluations and evaluations by their students. *Journal of Educational Psychology, 71*, 149-160.
- McKeachie, W. J. (1979). Student ratings of faculty: A reprise. *AAUP Bulletin, Academe, 65*, 384-397.
- Meier, R. S. and Feldhusen, J. F. (1979). Another look at Dr. Fox: Effect of stated purpose for evaluation, lecturer expressiveness, and density of lecture content on student ratings. *Journal of Educational Psychology, 71*, 339-345.
- Meredith, G. M. (1982). Grade-related attitude correlates of instructor/course satisfaction among college students. *Psychological Reports, 50*, 1142.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Orpen, C. (1980). Student evaluation of lecturers as an indicator of instructional quality: A validity study. *Journal of Educational Research, 74*, 5-7.
- Overall, J. U. and Marsh, H. W. (1980). Students' evaluations of instruction: A longitudinal study. *Journal of Educational Psychology, 72*, 321-325.
- Peterson, C. and Cooper, S. (1980). Teacher evaluation by graded and ungraded students. *Journal of Educational Psychology, 72*, 682-685.
- Smith, P. L. (1979). The generalizability of student ratings of courses: Asking the right questions. *Journal of Educational Measurement, 16*, 77-87.
- Thurston, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Vasta, R. and Sarmiento, R. F. (1979). Liberal grading improves evaluations but not performance. *Journal of Educational Research, 71*, 207-211.
- Whitely, S. E. and Doyle, K. O. (1979). Validity and generalizability of student ratings from between-classes and within-class data. *Journal of Educational Psychology, 71*, 117-124.
- Wotruba, T. R. and Wright, P. L. (1975). How to develop a teacher-rating instrument: A research approach. *Journal of Higher Education, 46*, 653-663.