

EFFECTS OF STUDENT ANONYMITY-NONANONYMITY ON THE FACTOR STRUCTURE OF A TEACHER RATING FORM

Royce R. Ronning and Ursula R. Walsh, *University of Nebraska, Lincoln*

.....
Factor analysis of an instructor rating form administered to three successive student and teacher populations revealed a reasonably consistent factor structure across analyses. In one of the three administrations, students were asked to sign the evaluation form; in this case, substantial changes in proportions of common variance appeared for the first two factors when comparing anonymous versus nonanonymous conditions. Results are discussed in terms of methods for use of student ratings to improve instruction.
.....

Key words: teacher ratings; factor analysis

The use of student evaluations of instructors as one index of teacher instructional effectiveness continues despite criticisms of the validity of student ratings (see Bryant, 1967; Rodin and Rodin, 1972, as examples). In view of their continued use, it is important that such evaluations be carefully examined not only for the nature of the content of the evaluation forms, but also for the circumstances under which the data are collected from students.

A considerable body of research (Werdell, 1967; Eble, 1970; Deshpande et al., 1970) examines, primarily through use of factor analytic techniques, student completed instructor evaluation forms for the purpose of identifying or clarifying underlying "structure." As a result of these studies, it appears that while a considerable number of independently developed evaluation forms exist, the various factor

Address reprint requests to Dr. R. R. Ronning, TC 20, Department of Educational Psychology and Measurements, University of Nebraska, Lincoln, NB 68588.

analytic solutions exhibit substantial overlap in factor content. Thus in their review of research on college teaching, Trent and Cohen (1973) have identified five major factors common to a number of factor analytic studies: (1) clarity of organization, interpretation and explanation; (2) encouragement of class discussion and the presentation of diverse points of view; (3) stimulation of students' interests, motivation and thinking; (4) manifestation of attentiveness to and interest in students; (5) manifestation of enthusiasm. Since each evaluation form is presumably designed to meet local conditions and since choice of wording and evaluation instrument format differ from instrument to instrument, total factor congruence is extremely unlikely. The extent to which similar factors appear in studies conducted on a wide variety of college and university campuses suggests progress toward the goal of ascertaining the bases for student evaluations of instruction. On the other hand, factor-analytic techniques cannot, of course, determine whether or not the domain of possible factors related to instructional effectiveness has been exhausted.

The extensive research into evaluation instruments has not been accompanied by systematic investigation of the conditions under which student evaluations are secured. In the authors' experience, wide variation in practice exists ranging from asking students to provide free response evaluations on the reverse of a signed course final examination to elaborate attempts to prepare formal evaluation instruments from which student responses are solicited under carefully controlled conditions guaranteeing student anonymity. Costin et al. (1971) assert that student anonymity is mandatory under conditions where ratings are obtained prior to receipt of final grade. However, they provide no data supporting the assertion. Recently, Abrami et al. (1976), in the context of a larger study, examined student anonymity on instructor rating forms as a main effect, finding it to result in no significant differences even though the ratings were elicited prior to final grading.

The possibility exists, however, that more subtle changes in student responses may occur than those subsumed under the category of overall ratings. Anonymity effects might be better evaluated by considering them as sources of information which shape the manner in which individual items are answered. Thus it is possible that the factor structure of an evaluation instrument might be modified without affecting the overall ratings.

The purpose of this study, then, is to evaluate the effect of one variable, student anonymity or nonanonymity, on the internal factor structure of an undergraduate instructor evaluation instrument.

METHOD

A required undergraduate course in educational psychology at the University of Nebraska, Lincoln, has used the same 28-item instructor and course rating form for a number of years. The form was originally constructed on a rational basis after examining a fairly large number of other instructor and course rating forms as well as the framework of the particular course in question. The form has been identical for each year with the exception of changes in the titles of textbooks used for one item of the form. Students respond by placing a checkmark along a line which ranges from an adjective description of superior instructor or course performance to ineffective instructor or course performance for items such as "instructor self-confidence" where the extremes range from "always sure of himself, meets difficulties with poise" to "hesitant, timid, uncertain." The scale consists of nine points.¹

The spring semester, 1973, evaluations were factor analyzed in order to allow the course supervisor and the course instructors to obtain a clearer sense of the bases for student evaluations. The factor analysis yielded a fairly clear solution resulting in four factors: Popular Teacher—reflecting characteristics of the instructor which promoted effective relationships with students; Course Evaluation—reflecting an evaluation of the course in terms of student enjoyment, extent of learning, and willingness to recommend the course to other students; Effective Teacher—reflecting such qualities of teacher and student interaction as clarity of teacher explanations, quality of answers to student questions, as well as items reflecting overall evaluations of the instructor; Teacher Fairness—reflecting only three items, fairness in grading, clarity of assignments and "valid" examinations. These factors bear some similarity, but are certainly not identical to, the five factors listed above from the Trent and Cohen review.

In 1974 the authors conducted an examination of student performance in the same course during which measures of grade-point average, ability, etc. as well as instructor evaluations were obtained from each student. As a consequence students were asked to sign the evaluation form so that the evaluation data could be individually coded along with other relevant material. The 1974 evaluations were distributed and collected by an individual other than the course instructor and students were assured that only the class averages would be shown to the instructor and that the signed forms would never be made available to the instructor. In spite of these assurances, eight students (of 230) refused to sign the forms and were permitted to turn them in unsigned.

In the spring semester, 1975, the evaluations were again collected

anonymously in a manner identical to the 1973 data collection. No other student data were gathered in either the 1973 or 1975 data collection periods.

RESULTS

Available for analysis then were the instructor and course evaluations of six teachers based on 229 students from 1973, of seven teachers based on 222 students in 1974, and of seven teachers based on 218 students in 1975. Each set of evaluations was factor analyzed by the principal factor method (with iterations) followed by varimax rotation. No limit was set on the number of factors to be extracted for any of the analyses. In each case six factors were extracted. All factors reported had eigen values greater than one with the exceptions of factor 4 in 1973 and 1975 and factor 3 in 1975. The consistency of structure for these factors across all three years is presented as justification for their retention. Four factors from each analysis are reported and interpreted. In each analysis the first factor extracted accounted for approximately 70% of the common variance, with each of the remaining factors contributing from about 4% to 11% of the common variance (see Table I). Following the factor analyses, the writers attempted to "match" the factors visually across the three analyses by examination of items with loadings of 0.30 or above. Results are reported in Table I. In the judgment of the authors, the match was reasonably successful.

The crucial point of Table I is the difference between the 1974 common factor variance estimates for the largest factor and those for 1973 and 1975. The most dramatic shift is the reversal of the factors labeled "Popular Teacher" and "Course Evaluation." Thus, in 1973 and 1975 "Popular Teacher" was the first factor extracted and accounted for 68% and 72%, respectively, of the common variance; however, in 1974, "Popular Teacher" was the second factor extracted and accounted for only 11% of the common variance. On the other hand, in the 1973 and 1975 analyses, "Course Evaluation," the second factor extracted, accounted for 11% and 8%, respectively, of the common factor variance. In the 1974 analysis "Course Evaluation" was the first factor extracted and accounted for 64% of the common factor variance. Note that while the same items receive high loadings on each of the first two factors across the three years, there is a very substantial shift in the proportions of common factor variance accounted for by the two factors between the 1974 and the 1973, 1975 analyses.

In an effort to provide further information about this change, means

TABLE I. Factor Loadings by Item by Year^a

Item/Loading	Popular Teacher			Course Eval.					Effective Teacher			Fair Teacher			
	1973	1974	1975	1973	1974	1975	1973	1974	1975	1973	1974	1975	1973	1974	1975
	One	Two	One	Two	One	Two	Three	Three	Three	Four	Four	Three	Four	Four	Three
1	(23)	(29)	74	(14)	36	(20)	75	51	(-03)						
2	(19)	42	66	(16)	46	(19)	65	(29)	(01)						
3	36	(27)	58	(11)	33	(16)	54	33	(-01)						
4	48	(26)	73				47	74	(19)						
5	30	(17)	65				67	80	33						
6	32	(29)	58				56	69	41						
7							38	61	(28)				(25)	31	(18)
8	(18)	50	58	36	30	37	64	34	(12)						
9	60	70	40												
10	44	51	64				45	35	(23)						
11	75	64	37				39	(26)	38				(03)	(04)	52
12	63	73	(23)										(22)	(09)	70
13	60	59	49										(20)	36	(08)
14	53	56	52												
15	40	54	68	32	40	52	73	56	34						
16	40	50	68	38	38	46	69	58	32						

TABLE I. (Continued)

Item/Loading	Popular Teacher			Course Eval.			Effective Teacher			Fair Teacher		
	1973	1974	1975	1973	1974	1975	1973	1974	1975	1973	1974	1975
	One	Two	One	Two	One	Two	Three	Three	Five	Four	Four	Three
17				38	(22)	(18)				(09)	55	(08)
18				(23)	32	33	33	46	(00)	(27)	51	36
19				(22)	(22)	33	41	38	(11)	62	52	(24)
20	44	31	(15)	42	35	(13)	(13)	34	(07)	70	40	54
21							(-01)	31	(05)	(02)	37	(16)
22												
23												
24	(20)	(20)	46	68	79	65	55	(27)	(21)			
25	(14)	(14)	32	82	87	74	32	(20)	(20)			
26				77	80	75	33	(10)	(10)			
27												
28				73	80	-35	36	(22)	(13)			
% Variance explained	68%	11%	72%	11%	64%	8%	7%	9%	4%	6%	7%	6.5%

^aDecimals omitted. Only loadings above 0.30 are presented. When an item has loadings of 0.30 or higher on only one analysis, the other loadings are presented in parentheses.

and standard deviations for items with the highest and most-consistent loadings on each factor are presented in Table II. Note that for the "Popular Teacher" factor, the item means average about 0.4 point higher in 1974 than in 1973, and 0.3 higher in 1974 than in 1975. Furthermore for each item on this factor the standard deviation is lower than for the same item in 1973 or 1975. For none of the other factors is such a trend apparent.

DISCUSSION

The same factor analytic procedure was applied using the same program for each of the three samples. The samples appear to be of reasonable size for factor stability, and inspection suggests that the factor structure as such was remarkably consistent across the three years considering that the results are based on three entirely different student and teacher populations. On a post-hoc basis then, the writers reason that the act of signing the evaluation instrument (even with the assurance of anonymity for the actual instructor) led the 1974 students to change the "emphasis" within the evaluations to the extent that the major component apparently shifts from focus on characteristics of the instructor to an overall evaluation of the course. Table II, then, provides some evidence to support an hypothesis that under signed (nonanonymous) evaluation conditions, students rate items evaluating personal characteristics of the instructor more highly and with less variability than under entirely anonymous conditions. The factor analytic technique is, of course, sensitive to item score variability and the change in proportions of common factor variance shown in Table I appears to reflect that sensitivity. Of interest, of course, is the rather substantial size of the change in common factor variance estimates as well as the fit of the findings with one's intuitive prediction of student performance under the nonanonymous conditions.

Costin et al. (1971) suggest that "The context in which student ratings of teachers are obtained is a matter of great importance. Clearly a guarantee of anonymity or immunity would seem to be mandatory, especially when ratings are administered (as is typical) prior to final grading." The present study provides mixed support for the statement. On the one hand, signing the course evaluations (even with the guarantee that the instructor would not see the signed form) did, apparently, shift the pattern of ratings. On the other hand the factor structure (ignoring common variance estimates) of the rating instrument remained reasonably consistent in the sense that the same items loaded on the same factors. The consistency of factor structure is consistent with the general results of Abrami et al. (1971), who found no

TABLE II. \bar{X} and σ for Each Major Item by Factor

Item No.	1973		1974		1975	
	\bar{X}	σ	\bar{X}	σ	\bar{X}	σ
9	7.56	1.37	7.96	1.23	7.69	1.88
10	7.48	1.37	7.44	1.31	7.45	1.47
11	7.59	1.53	7.99	1.40	7.78	1.59
13	7.46	1.44	8.06	1.22	7.31	1.78
14	7.02	1.54	7.57	1.33	7.30	1.54
Popular teacher	7.41		7.80		7.51	
8	6.62	1.69	6.58	1.51	6.84	1.51
21	6.29	2.23	6.47	2.33	6.76	2.21
24	5.71	1.86	5.87	1.80	5.96	1.82
25	6.11	1.76	5.80	1.81	6.02	2.04
26	5.56	1.98	5.28	1.98	5.71	2.16
Course evaluation	6.06		6.00		6.26	
5	6.89	1.63	6.57	1.69	6.84	1.79
6	7.20	1.38	7.24	1.40	7.34	1.67
7	7.36	1.44	7.11	1.52	7.20	1.87
15	6.50	1.97	6.54	2.06	6.72	2.05
16	6.61	1.95	6.81	2.03	6.97	2.18
Effective teacher	6.91		6.85		7.01	
18	6.99	1.72	6.78	1.70	7.40	1.53
19	7.16	1.51	6.76	1.60	6.95	1.82
20	7.80	1.31	7.70	1.32	8.04	1.42
Fair teacher	7.32		7.08		7.46	

difference in instructor ratings in an anonymous versus nonanonymous teacher evaluation situation. The findings of this study suggest, however, that any such studies should be accompanied by techniques such as those used here to search for changes in the relative contribution of the dimensions underlying the evaluation.

The four factors identified in Table I suggest two additional conclusions. Apparently students can and do consistently discriminate between instructor personal characteristics and interactive skills which make them well liked, as contrasted to characteristics which make instructors effective in the sense of providing clear explanations adjusted appropriately to the level of the class. An individual instructor may, of course, be effective, and thus well-rated, on both factors; however, the point to be made is that students apparently see these

two functions as discriminably different. Furthermore, fairness in grading, as well as in other course requirements, appears to be largely independent of both instructor factors. These findings suggest that students do not rate simply on the basis of some overall instructor bias. As a consequence, it seems reasonable to believe that rating forms based on these factors might well serve as a basis for improving instruction. Such an examination may reveal information permitting recommendations for instructor behavior change not only on personal warmth dimensions but also for effecting clearer communication with students as well as an assessment of the means used to evaluate student mastery of course material.

FOOTNOTE

¹Copies of the evaluation instrument are available from the senior author on request.

REFERENCES

- Abrami, P. C., Leventhal, L., Perry, R. P., and Breen, L. J. (1976). Course evaluation: How? *Journal of Educational Psychology* 68:300-304.
- Bryant, P. J. (1967). By their fruits ye shall know them. *Journal of Higher Education* 38:326-330.
- Costin, F., Greenough, W., and Menges, R. (1971). Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research* 41:511-535.
- Deshpande, A. S., Webb, S. C., and Marks, E. (1970). Student perceptions of engineering instructor behaviors and their relationships to the evaluation of instructors and courses. *American Educational Research Journal* 7:289-305.
- Eble, K. E. (1970). "The Recognition and Evaluation of Teaching." Salt Lake City, Utah: The Project to Improve College Teaching.
- Rodin, B., and Rodin, M. (1972). Student evaluations of teachers. *Science* 177:1164-1166.
- Trent, J. W., and Cohen, A. M. (1973). Research on teaching in higher education. In R. Travers (ed.), *Second Handbook of Research on Teaching*. New York: Rand McNally.
- Werdell, P. R. (1967). *Course and Teacher Evaluation*. (2nd ed.). Washington, D.C.: United States National Student Association.