DOCUMENT RESUME

ED 222 154 HE 015 597

AUTHOR Van Fleet, David D.; Chamberlain, Howard
TITLE Persistent Ratee Contaminants in Performance Appraisal.
PUB DATE 82
NOTE 13p.; Paper presented at the Annual Meeting of the American Psychological Association (Washington, DC, August 1982).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Evaluation Methods; *Faculty Evaluation; Higher Education; Questionnaires; *Rating Scales; Research Problems; *Student Evaluation of Teacher Performance; Teacher Effectiveness; *Validity

ABSTRACT

The hypothesis that conventional approaches to evaluating contaminants in performance appraisal overlook important individual ratee effects was examined. A rating form was developed that consisted of the following dimensions and behaviors: warmth; guided discourse or indirect teaching methods; control of subject matter; enthusiasm and reinforcing; organizing and managing; presenting and explaining; evaluating; and advising and counseling. Administration of the form to evaluate 23 instructors resulted in approximately 1,500 observations per semester. The reliability of the form and its factor stability were assessed, and possible contaminants were checked to assure that the evaluations were more likely to result from the instructor's performance than from student or course factors. It was found that 8.6 percent of the instructor ratees had persistent and significant contaminants associated with their evaluations; a looser definition of "persistent" pushes the figure to 34.7 percent. It is suggested that the evaluations may not be assessing performance accurately because of ratee contaminants, including expected grade in the course, the time at which the course begins, the time and effort required of the student, and the student's major. These contaminants occurred in spite of the fact that the instrument was found to have face validity, factor stability, and internal consistency. It is proposed that adjustments could be made on an individual basis and only for those contaminants that are persistent for each instructor. However, what is needed is a practical decision rule that would permit users of such evaluations to make necessary adjustments in the appraisals to correct for such persistent effects. Interactions among contaminants should also be addressed. (SW)

PERSISTENT RATEE CONTAMINANTS IN PERFORMANCE APPRAISAL

Subject Index: 18

by

David D. Van Fleet
Department of Management
Texas A&M University
College Station, TX 77843
(713) 845-3233

and

Howard Chamberlain
Department of Management
Texas A&M University
College Station, TX 77843

Sponsored by: Lyle F. Schoenfeldt
APA and Division 5 Member

Running Head: Research on Performance Appraisal

PERSISTENT RATEE CONTAMINANTS IN PERFORMANCE APPRAISAL

Abstract

This paper examines the hypothesis that conventional approaches to evaluating contaminants in performance appraisal overlook important individual ratee effects. "Important" in that they may be different than those identified for the total set of ratees and in that they persist over time. A form was developed and applied to 23 instructors, for each of four semesters resulting in at least 1500 observations per semester. The existence of persistent ratee contaminants is demonstrated. Further, the contaminant set identified for the individual ratees is not identical to that identified by the total ratee set. Implications are also briefly discussed.

PERSISTENT RATEE CONTAMINANTS IN PERFORMANCE APPRAISAL

Good or effective teaching is universally agreed upon as important and desirable. Yet, despite an increasingly large body of literature on the subject (see Selected Bibliography for examples), little is known precisely about what a good or effective teacher is. Efforts to isolate the essential differences between good and poor teachers are numerous, especially at the elementary and secondary school levels, but, while progress has been made, much remains to be done. It is not yet very clear which portions of a teacher's behavior are essential for learning and which are essential for student satisfaction with the learning process, for instance.

Colleges and universities have developed and/or are developing performance appraisal instruments to provide some student involvement in the evaluation of teaching. These instruments are being used increasingly by administrators as at least one source of information upon which to make personnel decisions. Departmental or subject matter specific instruments are rarely developed despite the obvious problems of aggregation and applicability which result from using a university-wide instrument. But administrative actions based on even carefully developed instruments may be in error due to the existence of persistent ratee contaminants. The purposes of this paper are first to demonstrate the existence of such persistent effects and then to briefly discuss their implications.

The Form

The particular form used in this research was developed as a departmental or subject matter specific form by a faculty committee consisting of a lawyer, an expert in performance evaluation in both the private and

public sectors, and a faculty member familiar with basic research in teaching evaluation. This committee examined the literature and several existing instruments and decided to strive for relevancy in the items used. Therefore, items which the literature suggested as being non-relevant were immediately dropped from consideration--dress, hair length, sex, and the like. A tentative list of items covering "dimensions" and "behaviors" identified in the literature was developed. The "dimensions" and "behaviors" were: (1) warmth; (2) guided discourse or indirect method of teaching; (3) control or grasp of subject matter; (4) enthusiasm, motivating and reinforcing; (5) organizing, coordinating and managing; (6) presenting, explaining and demonstrating; (7) evaluating; and (8) advising and counseling. That tentative list was submitted to the departmental faculty and a form finalized following faculty review.

The form was used in all sections of all courses taught by the department each regular semester for two years. In each semester, there were 1500 or more observations obtained (student forms, not separate students, as a single student may have had more than one instructor). A total of 23 instructors were evaluated all four semesters; to assure comparability, only these 23, were used in the analysis although more than 30 taught each of the semesters.

A split-half reliability coefficient (Spearman-Brown) was calculated for each semester. The results were: 0.83 (N=2007); 0.84 (N=1627); 0.84 (N=1798); and 0.84 (N=1499). Factor stability was also found to exist as can be readily seen in Table 1. These data suggest that the internal consistency and stability of the form was acceptable or even excellent.

To check for possible contamination, several items of student background and perceptions about the course were obtained and were correlated

TABLE 1

FOUR SEMESTER FACTOR ANALYSIS
(Items Grouped by Highest Loading on Rotated Factor Matrix)
(Principal Components Method; Varimax Rotation)
(Each * Represents One Semester)

	Factor					
Item	Openness	Interest in Students	Testing	Mechanics	View of Material	Main Point of Item where S refers to students and I to instructor
5	* * * *					S's feel free to ask questions
9	* * * *					Instructor asked challenging questions
14	* * * *					Instructor open to other views
22	* * * *					Instructor used Socratic method
6	*	* * *				Instructor provided guidance
2		* * * *				I met with S outside of class
15		* * * *				I personally interested in S
3		*	*	*		examination feedback useful
10			* * * *			exam questions were clear
16			* * * *			I fair in grading examinations
17			* * * *			exams were a good sample of material
20			*	* * *		course requirements and grading clear
4				* * *	*	Instructor organized
8				* * *	*	course objectives clear
11				* * *	*	Instructor was well prepared
19				* * *	*	I stressed important material
21				* * *	*	I accomplished objectives of course
25				* * *	*	material covered at satisfactory pace
12				* *	* *	I aroused S interest in material
13				* *	* *	concepts clearly presented
1					* * * *	Instructor knew material
7					* * * *	I displayed interest in material
18					* * * *	I used examples to clarify material
23					* * * *	I enthusiastic about material
24					* * * *	I presented information not in text

NOTE: Using eigenvalues of 1.0 or greater, five factors were obtained for three semesters and four factors for the fourth semester. For that fourth semester, the "Mechanics" factor was not separated from the "View of Material" factor as it was in the other three semesters.

with the overall ratings which were given to the instructors. The data presented in Table 2 demonstrate that while several highly significant contamination effects are present, they are quite small and, indeed, singly would account for extremely small proportions of variance in the overall ratings. These data, too, suggest that the form used was acceptable or even better.

Ratee Contaminants

Thus far what has been done is rather conventional (with the exception of four replications of factor analysis) for the evaluation of teaching. A faculty committee was used to assure face validity and acceptance in the form developed. The reliability of the form and its factor stability were assessed to assure the internal consistency and stability of the form. Possible contaminants were checked to assure that the evaluations received were more likely to be the result of the performance of the instructor than characteristics of the students or the course. In this instance, all indicators were that the form was acceptable but that slight contaminants appeared to exist--the grade which the student expects to receive in the course, the sex of the instructor, and the time of day at which the course begins.

But what if one or more of the instructors performed in such a way as to evoke a persistent pattern of response from students with regard to the evaluation--a pattern reflecting not performance but some characteristic of the student? What if an instructor constantly made sexist remarks so that females underrated the instructor while males were mixed in their reactions? Such ratee specific contaminants might go undetected using a conventional approach to evaluate the form such as outlined above. The usual assumption seems to be that such effects might exist but would not persist. Therefore,

TABLE 2

CONTAMINATION ANALYSIS

Background Factor	Semester 1	2	3	4
Major	+0.0050	-0.0080	-0.0548	-0.0032
Classification	+0.0290	-0.1176***	-0.0253	+0.0191
Time Course Begins	+0.0631*	+0.0008	+0.0745*	+0.1362***
Reason for Selecting Course	-0.0049	+0.0045	-0.0482	-0.0308
Career Plans	+0.0154	-0.0409	-0.0089	-0.0346
Sex[a]	+0.0294	+0.0227	+0.0612	+0.0290
Sex of Instructor[a]	+0.1057***	+0.1150***	+0.1160***	+0.1216***
Match between Student Sex and Instructor Sex	+0.0686**	+0.0903***	+0.1148***	+0.0397
Overall Grade Point Average	-0.0294	+0.0778**	+0.0636**	-0.0281
Expected Grade in Course	+0.1363****	+0.1905***	+0.2304***	+0.1248***
Difficulty of Course	-0.0395	-0.0918***	-0.0875***	+0.0266
Time and Effort Required in Course	-0.0107	-0.1035***	-0.1439***	+0.0244

* $p \leq 0.01$

** $p \leq 0.001$

*** $p \leq 0.0001$

[a]Sex was coded 1 for males and 2 for females.

an examination of those effects by instructor (N=23) for each semester (N=4) was performed. The results are summarized in Table 3.

As can be seen from the data in Table 3, 8.6 percent of the instructor ratees have persistent contaminants associated with their evaluations where persistent is defined to mean significant ($p \leq 0.10$) in all four semesters. If a slightly looser definition of persistent is used--significant in three of the four semesters, this increases to 34.7 percent, over a third of the instructors! This means, of course, that for those ratees, the evaluations may not be assessing performance accurately.

One of the twelve possible contaminants is involved for six of the eight faculty--expected grade in course--and the correlation is as high as 0.59 for one of the instructors in one semester. The other persistent contaminants are the time at which the course begins, the time and effort required of the student in the course, and the student's major. These latter two were not identified as persistent when the total data set was analyzed. Further, the sex of the instructor, which had been identified as a persistent contaminant for the total data set, does not appear as one on an instructor-by-instructor basis.

Conclusions

These data clearly demonstrate the existence of persistent ratee contaminants even for an instrument which is acceptable to those for whom it is designed to be applied and which appears to be reasonably good from a psychometric standpoint. This means that, even if conventional approaches are used to assure good evaluations are being conducted, some ratees may be receiving improper evaluations--higher or lower than their performance alone would suggest.

What can be done? First, even if the total data set is used in a

TABLE 3

RATEE CONTAMINATION ANALYSIS

Ratee	1	2	3	4
1	j		k	f,h,j
2	a,b,i			
3	(c),(j)	(j),l	b,(c),f,h,(j)	a,(c),i,k
4	f,h,j	e,f,h,j	b,c,i,k	k
5		i,j		
6		a,f,h,l	c	i
7	b,(c)	(c),k	(c),j	j
8	c,f,h,i,(j),k,l	c,(j),l	(j)	e,(j)
9	(j),k	e,f,h,(j)	f,h,(j)	k,l
10	a,b,f,h,i	a,b,d	c,j	d
11	j	a,f,h		d,i
12	b,c	f,h	l	a,c,d,h,j
13	l	b,j,k		
14	f,h,(j),(l)		f,h,(j),(l)	b,i,(j),(l)
15	a,i	b	c,i	d,f,h,k,l
16	f,h,(j),k,l	f,h,(j)	c,(j),l	(j),k
17	(a),c,d,i,(j),k	(a),f,h	(a),b,d,f,h,(j),l	(j)
18	b,c,k,l	h,j	i,j	a,b,c,d,i,k,l
19		j	b,c,f,h,j	b,c,e,f,h,l
20.	a	j	k	d
21		c,i,j,l	b,c,j	l
22	e,(j)	a,c,(j),k	a,b,c,d,(j)	
23	a,d,i			f,h,j,k

NOTE: Only those contaminants significant at the 0.10 level are shown.

a = major
b = classification
c = time course begins
d = reason for selecting course
e = career plans
f = sex of student (1=male;2=female)
g = sex of instructor (1=m;2=f)
h = match between student sex and instructor sex
i = overall grade point ave.
j = expected grade in course
k = difficulty of course
l = time and effort required in course

( ) denotes persistent contaminant for particular ratee/instructor.

regression model to adjust for possible contaminants, proper results may not be obtained. As indicated by these data, contaminants which appear significant for the total data set may not be persistent ones for ratees and some which are persistent for ratees may not show up for the total data set. Second, the contaminants which emerge with each administration of an instrument are likely both to involve more ratees than those for whom persistent contaminants exist and to involve more possible contaminants than actually will persist. This would seem to suggest that adjustments would have to be made on an individual basis and only for those contaminants which are persistent for each individual. This is clearly a monumental task and one which is not likely to be undertaken by many organizations.

What is needed is a practical decision rule which will permit users of such evaluations to make necessary adjustments in the appraisals to correct for such persistent effects without all of this effort, if possible. However, before that can be done the question of interactions among these contaminants must also be addressed. It is highly likely that "reason for choosing the course," "career plans," "classification," and "major" will display some interaction which may heighten or lessen the problem identified here. The next step in the research project being reported here is to examine those interactions and to move toward the establishment of a decision rule for treating the problem.

SELECTED BIBLIOGRAPHY

Brandenburg, D. and Remmers, P., "Rating scales for instructors," Educational Administration and Supervision, Vol. 13 (1927), 459-466+.

Breed, F. S. "Factors contributing to success in college teaching," Journal of Educational Research, Vol. 16 (1927), 247-253.

Campbell, D. T. and Fiske, D. W., "Convergent and discriminant validation by the multi-trait, multi-method matrix," Psychological Bulletin, Vol. 56 (1959), 81-105.

Carrier, N. A., et al., "Course evaluation: when?" Journal of Educational Psychology, Vol. 66 (1974), 609-613.

Champlin, C., "The preferred college professor," School and Society, Vol. 27 (1928), 25-37.

Charters, W. W., "The improvement of college teaching," School and Society, Vol. 13 (1921), 494-497.

Coats, W. D., et al., "Student perceptions of teachers: a factor analytic study," Journal of Educational Research, Vol. 65 (1972), 357-360.

Davenport, M. M., "Methods for evaluating good teaching," Journal of Animal Science, Vol. 24 (1965), 1209-1214.

Davis, C. O., "Our best teachers," School Review, Vol. 34 (1926), 754-759.

Elbe, K. E., "Improving college teaching," Phi Delta Kappan, Vol. 52 (1971), 283-285.

Frey, P. W., "Student ratings of teaching: validity of several rating factors," Science, Vol. 182 (1973), 83-85.

Gage, N. L. (ed.) Handbook of Research on Teaching (Chicago: Rand McNally, 1963).

Isaacson, R. L., et al., "Dimensions of student evaluations of teaching," Journal of Educational Psychology, Vol. 55 (1964), 344-351.

Jenkins, J. R. and Davis, S. L., "Influence of student behavior on teacher's self-evaluation," Journal of Educational Psychology, Vol. 60 (1969), 439-442.

Kent, L., "Student evaluation of teaching," The Educational Record, Vol. 47 (1966), 376-406.

Leventhal, D. F., et al., "Student evaluations of teacher behaviors as estimations of real-ideal discrepancies: a critique of teacher rating methods." Journal of Educational Psychology, Vol. 62 (1971), 104+.

Marks, E., "Individual differences in perceptions of the college environment," Journal of Educational Psychology, Vol. 61 (1970), 270-279.

Maslow, A. H. and Zimmerman, W., "College teaching ability, scholarly activity, and personality," Journal of Educational Psychology, Vol.

McFillen, J. M., "An analysis of factor congruency and subscale reliability for a course evaluation questionnaire," Proceedings of the American Institute for Decision Sciences, 1976.

Nash, H. B. and Bush, F. R., "To what extent do grades influence student ratings of instructors?" Journal of Educational Research, Vol. 21 (1930), 314-316.

Nunnally, J. C., et al., "Factored scales for measuring characteristics of college environments," Educational and Psychological Measurement, Vol. 23 (1963), 239-248.

Pambookain, H. S., "Initial level of student evaluation of instruction as a source of influence on instructor change after feedback," Journal of Educational Psychology, Vol. 66 (1974), 52-55.

Perry, R. P. et al., "Effect of prior teaching evaluations and lecture presentation on ratings of teaching performance," Journal of Educational Psychology, Vol. 66 (1974), 851-856.

Pressey, S. L., et al., "Research adventures in university teaching," School and Society, Vol. 20 (1924), 635-638.

Rees, R. D., "Dimensions of students' point of view in rating college teachers," Journal of Educational Psychology, Vol. 60 (1969), 476-482.

Rodin, M. and Rodin, B., "Student evaluations of teachers," Science, Vol. 177 (1972), 1164-1166.

Spencer, R. E. and Aleamoni, L. W., "A studnet course evaluation questionnaire," Journal of Educational Measurement, Vol. 7 (1970), 209-210.

Sullivan, A. M. and Skanes, G. R., "Validity of student evaluation of teaching and the characteristics of successful instructors," Journal of Educational Psychology, Vol. 66 (1974), 584-590.

Van Fleet, D. D., "Salary administration in higher education: a tentative plan," AAUP Bulletin, Vol. 58 (1972), 413-418.

Van Fleet, E., "Q-sort in teacher evaluation," The Delta Pi Epsilon Journal, Vol. 14 (1972), 1-17.

Veldman, D. J. and Peck, R. F., "Influences on pupil evaluations of student teachers," Journal of Educational Psychology, Vol. 60 (1969), 103-108.

Woodring, P., The higher learning in America: a reassessment (New York: McGraw-Hill, 1968).