DOCUMENT RESUME

ED 201 659 TM 810 252

AUTHOR Wolfe, Mary L.
TITLE A Causal Modeling Approach to the Analysis of Course Evaluation Data.
PUB DATE Apr 81
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Correlation; *Course Evaluation; *Course Objectives; Higher Education; Instructional Materials; Learning Activities; Masters Programs; Nursing Education; *Student Evaluation of Teacher Performance
IDENTIFIERS *Causal Models; *Likert Scales

ABSTRACT

Five causal models relating several aspects of end-of-term student evaluation of a graduate course in nursing research methods were proposed and tested empirically. The course evaluation form consisted of four Likert-type subscales, on which students rated the following aspects of the course: (1) the extent to which the course met its objectives; (2) the utility of various instructional resources and activities in meeting course objectives (e.g., reading materials, quizzes, examinations); (3) the effectiveness of a number of specific course policies (e.g., open book examinations and quizzes) in promoting learning; and (4) instructor effectiveness. For each model, a prediction was made in terms of vanishing partial or zero-order correlations. The most plausible models suggested that ratings of the extent to which course objectives were met, and of instructor effectiveness, were both linked with ratings of course policy effectiveness via ratings of the utility of instructional resources in meeting course objectives. Course policies may best be implemented by means of appropriate instructional resources and techniques; students will then rate the instructor as effective and regard course objectives as having been met. (Author/RL)

A CAUSAL MODELING APPROACH TO THE ANALYSIS OF COURSE EVALUATION DATA

Mary L. Wolfe
School of Nursing
University of Maryland
Baltimore, Maryland

Presented at the 1981 Annual Meeting of the American Educational Research Association, April 13-17, 1981, Los Angeles, California

A Causal Modeling Approach to the
Analysis of Course Evaluation Data

MARY L. WOLFE
School of Nursing
University of Maryland

Abstract

Five causal models relating several aspects of student evaluation of a graduate course in nursing research methods were proposed and tested empirically. The most plausible models suggested that ratings of the extent to which course objectives were met, and of instructor effectiveness, were both linked with ratings of course policy effectiveness *via* ratings of the utility of instructional resources in meeting course objectives. Course policies may best be implemented by means of appropriate instructional resources and techniques; students will then rate the instructor as effective and regard course objectives as having been met.

a) Objectives

The purpose of the present study was to test alternative causal models relating several aspects of nursing students' evaluations of a graduate course in nursing research methods.

b) Perspectives/Theoretical Framework

Although causality can never be directly inferred from correlational data, causal inferences may be made concerning the adequacy of specific causal models (Simon, 1957; Blalock, 1964; Duncan, 1975). The method for making such inferences involves the explicit definition of a finite set of variables, assumptions about the causal connections among those variables, and assumptions about the effect of outside variables upon the variables in the model. Models are then eliminated which make predictions which are inconsistent with the data. The conditions imposed on each model are the predictions expressed in terms of the correlation coefficients (either zero-order or partial) that should be obtained if the model is correct . These predictions are always in terms of vanishing partial or zero-order correlation coefficients (Nachmias and Nachmias, 1976).

In addition to the agreement between predicted and obtained correlation coefficients, the temporal order among the variables must be taken into account in establishing causation. In responding to an evaluation questionnaire at the end of a course, students are typically asked to give a summary rating of a variety of different aspects of the course -- objectives, materials, lectures, instructor effectiveness and so on. Although it may be difficult

to establish a temporal order among these several aspects of the course at the time of evaluation, it is quite reasonable to assume that some causal relationship exists between them. For instance, instructor competence, or lack of it, may influence the framing of course objectives. On the other hand, poor objectives, or inadequate instructional materials, may render an otherwise capable instructor ineffective in the eyes of his or her students. These causal connections may, in turn, be reflected in student ratings.

c) Methods.

Data from student evaluations of a masters' level course in nursing research methods were subjected to correlational analysis as follows. (See Appendix for a copy of the evaluation form, along with means and standard deviations of student ratings of each item). The course evaluation form consisted of four Likert-type subscales, on which students rated these aspects of the course: the extent to which the course met its objectives (variable X_1); the utility of various instructional resources and activities in meeting course objectives (e.g., reading materials, quizzes, examinations) (variable X_2); the effectiveness of a number of specific course policies (e.g., open book examinations and quizzes, team research projects, critiquing published research in class) in promoting learning (variable X_3); and instructor effectiveness (variable X_4). Each subscale consisted of between eight and eleven items. Subscale reliabilities (internal consistency as indexed by Cronbach's alpha coefficient) are shown in Table 1.

Table 1. Subscale Reliabilities (Cronbach's Alpha Coefficient)

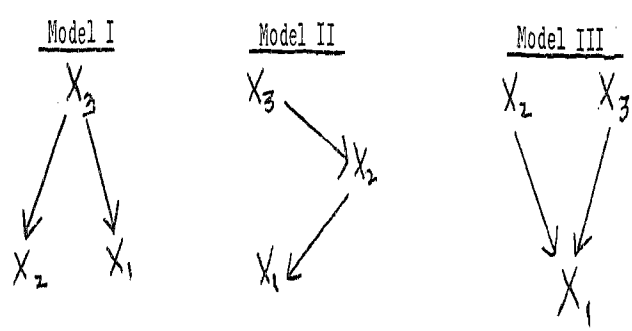
Subscale	Reliability Coefficient
1. Extent to which course met objectives	0.79
2. Utility of instructional resources and activities	0.83
3. Effectiveness of specific course policies	0.44
4. Instructor effectiveness	0.83

The matrix of intercorrelations among the four subscales is shown in Table 2 below.

Table 2. Intercorrelations Among Evaluation Form Subscales

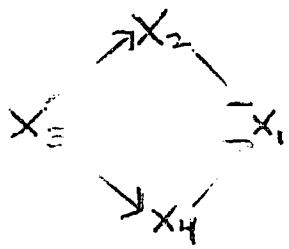
	X_1	X_2	X_3	X_4
X_1	-	.52	.35	.51
X_2	-	-	.53	.51
X_3	-	-	-	.47
X_4	-	-	-	-

Five causal models suggested by the work of Goldberg (1966) and Rehberg, Schafer and Sinclair (1970) were hypothesized and tested. The first three were three-variable models involving aspects of the course apart from instructor effectiveness.

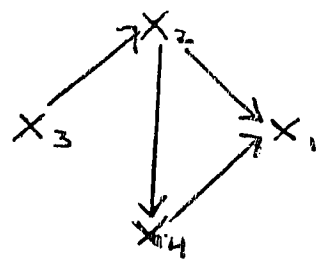


Two more models involving all four variables were tested next.

Model IV



Model V



d) Data Source

End-of-term student evaluations of a masters' level course in nursing research methods at a large eastern university were the data source for this study. The sample consisted of 10 students in three sections, all taught by the author. Confidentiality of responses was insured by the fact that students were asked not to identify themselves in any way.

e) Results

For each model, a prediction was made in terms of vanishing partial or zero-order correlations. Predictions and empirical results for the five models are shown in Table 3 on page 5. Of the three-variable models, Models I and III can be eliminated as implausible. Model II fits the empirical results fairly well, and hence, may be supported. Although neither Model IV nor Model V provides a close fit to the data, Model V appears to be the more plausible of the four-variable models. In summary, Model II suggests that students' perception of the utility of instructional resources in meeting course objectives intervenes between ratings of course policy effectiveness and of the extent to which course

Table 3. Predictions and Empirical Results for Models I - V

	Predictions	Empirical Results
Model I	$r_{12.3} = 0$	$r_{12.3} = .422$
Model II	$r_{13.2} = 0$	$r_{13.2} = .096$
Model III	$r_{23} = 0$	$r_{23} = .531$
Model IV	$r_{24.3} = 0$ $r_{13.24} = 0$	$r_{24.3} = .439$ $r_{13.24} = .027$
Model V	$r_{34.2} = 0$ $r_{13.24} = 0$	$r_{34.2} = .238$ $r_{13.24} = .027$

objectives are met. When a fourth variable (instructor effectiveness) is added to the model, as in Model V, the basic structure of Model II is still discernible -- but now, ratings of both instructor effectiveness and of the extent to which objectives are met are linked with ratings of course policy via the intervening variable, perception of the utility of instructional resources. It may be that course policy is best implemented by means of appropriately chosen instructional resources and techniques: in which case students will rate the instructor as effective and regard the course objectives as having been met.

f). Educational/Scientific Importance

A number of recent studies (e.g., Tetenbaum, 1975; Crittenden and Norr, 1975; Frey, Leonard and Beatty, 1975; Scott, 1975; Whitely and Doyle, 1976; and Marsh, 1977, 1980) have addressed the problem of the reliability and validity of student ratings of in-

struction. However, once these have been established, the question of practical use in the improvement of instruction remains. What are the relationships among the various aspects of student evaluation of instruction? What variables can or should be manipulated in order to improve the quality of instruction? In view of the difficulty of conducting true experiments in order to answer these questions, other, less direct approaches to exploring causal relationships must be considered. Although the present study involved a fairly small sample of students and relatively simple models, results suggest that the causal modeling approach may be of considerable value. A thoughtfully constructed causal model, empirically validated, can identify those variables in a given causal sequence whose manipulation is most likely to yield change in a desired direction. The almost universal practice of having students evaluate each course at its conclusion should provide a large and rich source of data for further exploration of the potential of this technique for assessing and improving instruction.

However, causal modeling should not be embraced with uncritical enthusiasm. Magoon (1978) has pointed out that enduring problems plague the application of causal modeling in a number of research settings. Not the least of these problems is the fact that researchers must always work with measuring tools which are less than perfectly reliable. The author of the present paper felt that it might be worthwhile to investigate, in a purely speculative spirit, whether significant changes would result in the interpretation of the results if all the correlations between the subscales of the evaluation tool were corrected for attenuation (Nunnally, 1978). Measurement

error -- unreliability -- makes correlations between measures less than they would be if measurement error were not present. From the theory of measurement error one may estimate how much correlations between true scores would be higher than those between fallible scores. If there were no error present, the correlation $\hat{r}_{12}$ between two measures would be given by

$$\hat{r}_{12} = \frac{r_{12}}{\sqrt{r_{11}}\sqrt{r_{22}}}$$

where r_{12} is the correlation between the fallible measures X_1 and X_2, and r_{11} and r_{22} are the reliabilities of X_1 and X_2, respectively. The equation above should really be regarded as an estimate rather than a correction -- it is an estimate of what the correlation between two measures would be if these measures were perfectly reliable. The results of applying the correction for attenuation to subscale intercorrelations are shown in Table 4 below.

Table 4. Intercorrelations Among Evaluation Form Subscales Corrected for Attenuation

	X_1	X_2	X_3	X_4
X_1	-	.64	.59	.63
X_2	-	-	.88	.70
X_3	-	-	-	.78
X_4	-	-	-	-

The zero-order and partial correlation coefficients in Table 3 were recalculated, using the values in Table 4. The results are shown in Table 5.

Table 5. Predictions and Empirical Results for Models I - V Based on Subscale Intercorrelations Corrected for Attenuation.

	Predictions	Empirical Results
Model I	$r_{12.3} = 0$	$r_{12.3} = .315$
Model II	$r_{13.2} = 0$	$r_{13.2} = .083$
Model III	$r_{23} = 0$	$r_{23} = .880$
Model IV	$r_{24.3} = 0$	$r_{24.3} = .060$
	$r_{13.24} = 0$	$r_{13.24} = -.091$
Model V	$r_{34.2} = 0$	$r_{34.2} = .480$
	$r_{13.24} = 0$	$r_{13.24} = -.091$

Of the three-variable models, Model II still appears to be the best supported by the data as before. However, of the four-variable models, Model IV clearly appears to be the more plausible, rather than Model V. Model IV suggests that student perceptions of the extent to which course objectives were met is linked with the effectiveness of specific course policies via ratings of instructor effectiveness and the utility of instructional resources and activities. The interpretation changes somewhat: course policy may best be implemented by means of appropriately chosen instructional resources applied by an effective instructor, in which case students will judge that the course objectives have been met.

As Nunnally points out (1978) there is some controversy regarding the correction for attenuation: one may deceive one's self into believing that a better correlation has been found than is actually warranted by the data. In addition, it is often

argued that the correction for attenuation provides a poor estimate of the correlation actually obtained between variables when they are made highly reliable.

If the preceding discussion does nothing else, it should suggest that if the causal modeling approach to analysis of course evaluation data is to be taken seriously as a guide to the improvement of instruction, a critical ingredient in the process is a course evaluation tool with highly reliable subscales.Efforts must be directed toward the development of such tools, since student course evaluations are becoming an ever more important consideration, not only in the improvement of instruction, but in the making of administrative decisions regarding promotion and tenure of faculty.

REFERENCES

Blalock, H.M. Causal Inference in Nonexperimental Research. Chapel Hill: University of North Carolin Press, 1964.

Crittenden, K.S. and Norr, J.L. Some remarks on "Student ratings: The validity problem". American Educational Research Journal, 1975, 12(4), 429-433.

Duncan, O.D. Introduction to Structural Equation Models. New York: Academic Press, 1975.

Frey, P.W., Leonard, D.W. and Beatty, W.W. Student ratings of instruction: Validation research. American Educational Research Journal, 1975, 12(4), 435-447.

Goldberg, A.S. Discerning a causal pattern among data on voting behavior. American Political Science Review, 1966, 60, 913-922.

Magoon, A.J. Path analysis: Application and explication. Evaluation and the Health Professions, 1978, 1(1), 94-99.

Marsh, H.W. The validity of students' evaluations: Classroom evaluations of instructors independently nominated as best and worst teachers by graduating seniors. American Educational Research Journal, 1977, 14(4), 441-447.

Marsh, H.W. The influence of student, course and instructor characteristics in evaluations of university teaching. American Educational Research Journal, 1980, 17(2), 219-237.

Nachmias, D. and Nachmias, C. Research Methods in the Social Sciences. New York: St. Martin's Press, 1976.

Nunnally, J.C. Psychometric Theory. (2nd edition). New York: McGraw-Hill, 1978.

Rehberg, R.A., Schafer, W.E. and Sinclair, J. Toward a temporal sequence of adolescent achievement variables. American Sociological Review, 1970, 35, 34-48.

Simon, H.A. Models of Man: Social and Rational. New York: Wiley, 1957.

Tetenbaum, T.J. The role of student needs and teacher orientation in student ratings of teachers. American Educational Research Journal, 1975, 12(4), 417-429.

Whitely, S.E. and Doyle, K.O. Implicit theories in student ratings. American Educational Research Journal, 1976, 13(4), 241-251.

APPENDIX

UNIVERSITY OF MARYLAND
SCHOOL OF NURSING

Nursing 701 - Course Evaluation

Section 1. On a scale (Likert) of 1 to 5, please assess how well the course met its objectives. 1 = not at all; 5 = completely met its objectives.

X	S.D.	
4.39	0.41	1. Applies the components of the research process in nursing.
4.46	0.65	2. Explains appropriate statistical procedures for estimating the reliability and validity of measurement.
3.99	0.91	3. Applies univariate and bivariate research designs and statistics.
4.13	0.80	4. Identifies selected multivariate research designs and statistics.
4.51	0.69	5. Produces or creates a research plan in collaboration with others.
4.52	0.64	6. Implements the research plan collaboratively.
4.51	0.73	7. Evaluates participation of self and collaborators as team members in carrying out the research plan.
4.22	0.60	8. Evaluates research studies employing univariate and bivariate designs and statistics.
4.02	0.84	9. Uses teletype and mini-computer for data analysis.

Section 2. Please assess, on a scale (Likert) of 1 to 5, the amount of helpfulness in meeting the course objectives of each of the following 1 = not helpful; 5 = very helpful.

3.09	1.14	10. Text A (required).
3.85	1.04	11. Text B (required).
2.63	1.58	12. Supplementary texts.
4.75	0.50	13. Handouts on content(i.e., Research Design, Correlation, Inferential Statistics).
3.89	0.85	14. Critiques of research studies in class.
3.98	0.89	15. Your critique of a research article.
4.37	0.80	16. Team research project.
3.90	0.97	17. Exams.
4.72	0.51	18. Instructor.

4.53	0.68	19. Class discussion and lectures
4.38	0.82	20. Quizzes

Section 3. Please assess on a scale (Likert) of 1 to 5, the usefulness of each of the following in terms of your learning and your being able to fulfill the course requirements. 1 = not at all useful; 5 = very useful.

3.77	1.16	21. Working by yourself rather than as a team on the critique of the research article.
3.99	1.21	22. Working as a research team rather than by yourself on the research project.
4.85	0.43	23. Taking open-book exams rather than closed-book exams.
4.82	0.51	24. Reviewing descriptive statistics the first day of class rather than never reviewing.
4.21	0.98	25. Submitting team progress reports rather than none.
4.59	0.64	26. Getting feedback from your instructor on progress reports rather than no feedback.
4.27	0.92	27. Critiquing assigned articles in class versus no class discussion.
3.70	1.34	28. Written reports rather than class time used for reporting of critiques of articles.

Section 4. Please assess on a scale (Likert) of 1 to 5, the following aspects regarding your instructor. 1 = poor; 5 = outstanding.

4.92	0.27	29. Preparation and knowledge of the subject.
4.80	0.42	30. Helpfulness to students.
4.37	0.70	31. Clarity in explaining concepts and answering questions.
4.55	0.62	32. Making the course interesting.
4.31	0.89	33. Feedback on team progress reports.
4.62	0.50	34. Lecture and discussion of course content.
4.34	0.76	35. Lecture and class discussion critiquing assigned research articles.
4.56	0.68	36. Availability for help outside of class.