

## A Small Study Assessing Teacher Effectiveness: Does Learning Last?

W. J. McKEACHIE, YI-GUANG LIN, AND  
CYNTHIA NEIGLER MENDELSON  
*University of Michigan*

Students of six classes who had previously participated in a larger study of teaching styles were tested a year after completion of the course. The purpose of the follow-up was to determine whether or not the prior battery of tests, including a student evaluation of instructor form, the Introductory Psychology Criteria Test, an Attitude Toward Psychology Scale, and a knowledge test, administered in a large group setting independent of those used for grades by instructors, would be positively related to student performance on comparable tests given a year later. The follow-up measures included items from the above measures plus questions regarding experiences and readings related to psychology in the year since the students' introductory course and two brief experiments which the students were to critique. Results indicated that the end of the semester measures of teaching effectiveness in terms of student performance and attitudes were positively related to similar responses obtained a year later.

Probably the oldest objection to student ratings is the comment, "I did not really appreciate some of my best teachers until sometime after the course had ended." Another common quote is, "Most of what a student learns and puts on a final examination is forgotten by the next week."

Is it true that end of course measures are not good indicators of longer lasting effects of teaching? This article reports the relationship between measures of teaching effectiveness gathered at the end of the course and measures administered a year later.

We have on previous occasions argued that course final examinations are not good measures of teaching effectiveness, even though in using them one makes the proper assumption that the ultimate criterion of teacher effectiveness is student learning. Unfortunately if we are comparing two methods of instruction or two teachers, final examination performance is an insensitive method of doing so. Students are so highly motivated for grades that they may well compensate for any deficiencies in the instruction they have received by extra preparation for the final examination. Thus, as reviews of teaching research indicate (e.g., McKeachie & Kulik, 1975), it is unusual to find experiments comparing teaching methods in which significant differences on a final examination are re-

Support for this research was provided by Grant GS-33054 from the National Science Foundation to W. J. McKeachie.

Reprint requests should be addressed to Dr. W. J. McKeachie, The Center for Research on Learning and Teaching, 109 E. Madison, Ann Arbor, MI 48109.

TABLE 1  
COMPARISONS OF END-OF-COURSE MEASURES WITH FOLLOW-UP MEASURES

Section	End-of-course up	Follow-up	N	Student ratings of instructor's general teaching ability		Interest in advanced psychology courses		Attitude toward psychology		Introductory psychology criteria test		
				End-of-course ratings	Follow-up ratings	Mean # desired at end-of-course	Mean # taken since end-of-course	End-of-course scores	Follow-up scores	End-of-course scores (48 items)	Follow-up scores (10 items)	
A	25	20		$\bar{X}$	4.12 <sup>1</sup> (2) <sup>2</sup>	3.90(1)	18.94(2)	2.05(3)	75.11(2)	75.20(1)	24.79(6)	4.55(5)
				SD	.86	1.16	12.15	2.68	9.25	10.68	6.28	1.85
B	23	16		$\bar{X}$	4.18(1)	3.62(2)	10.48(6)	.63(6)	73.00(5)	70.69(5)	30.50(1)	5.81(2)
				SD	.88	.96	6.99	.89	7.10	9.67	5.33	1.90
C	20	14		$\bar{X}$	4.08(3)	3.50(3)	22.00(1)	2.20(2)	77.18(1)	73.07(2)	26.00(4)	5.36(4)
				SD	1.00	.85	12.21	2.22	10.66	8.72	6.34	2.76
D	21	14		$\bar{X}$	3.65(4)	3.28(4)	12.18(5)	2.00(4)	70.47(6)	71.79(3)	27.35(3)	5.71(3)
				SD	.78	1.27	9.18	2.45	10.01	9.08	6.69	2.09
E	15	11		$\bar{X}$	3.35(5)	3.09(5)	17.07(3)	2.45(1)	74.00(3)	71.09(4)	27.62(2)	6.18(1)
				SD	.61	.83	11.94	3.27	12.30	10.32	6.27	1.47
F	21	17		$\bar{X}$	3.18(6)	2.82(6)	12.94(4)	.88(5)	73.67(4)	69.59(6)	25.39(5)	4.41(6)
				SD	1.01	1.13	8.97	.99	7.24	9.24	4.38	1.94

<sup>1</sup> Scale: 1 = Poor; 5 = Excellent.

<sup>2</sup> The section's rank order value for each measure is in parentheses following the mean score. Although the "Interest" measure is positively skewed; the rho using medians was of the same magnitude as that using the means.

vealed, except in cases in which the final examination has been contaminated by the inclusion of items on which only one group of students has had previous practice.

To avoid this problem, we have in our more recent research attempted to administer our measures of student learning in a setting independent of those used for grades by the instructor. In the present study, large group testing sessions were arranged for several nights during the two-week period prior to the end of classes and students were allowed to choose the evening most convenient for them.

The purpose of the present study was to determine whether or not tests administered in such a setting would be valid measures of teaching effectiveness as determined by student performance on comparable tests given a year after completion of the course.

## METHOD

### *Sample*

The sample consisted of six teachers of introductory psychology courses at The University of Michigan. The teachers were advanced graduate students working toward the PhD degree. These teachers and the students in their classes had volunteered to participate in a research study involving classroom observation of teaching styles (McKeachie, Lin, Daugherty, Moffett, Walz, Loomis, Neigler, Kerr, & Mann, 1974).

### *End of Course Measures*

(1) Student perception of teaching and teacher: The student rating form used in this study consisted of 32 items derived from the Michigan form described by Isaacson, McKeachie, Milholland, Lin, Hoffeller, Baerwaldt, & Zinn (1964).

(2) Introductory Psychology Criteria Test: The Introductory Psychology Criteria Test (Milholland, 1964) is a multiple-choice test designed to measure six levels of cognitive outcomes of an introductory psychology course: interpretation, application, analysis of elements, analysis of relationships, derivation of abstract relations, and judgment by external criteria. Forty-eight items were chosen from the previous forms of the test for this study.

(3) Attitude Toward Psychology Scale (Carrier, 1966): Twenty Likert-type items were administered to measure student attitude toward psychology.

(4) Advanced Courses: Students were asked to check which of the undergraduate psychology courses offered at the University of Michigan they would like to take.

### *Testing*

Tests were administered as part of a larger battery of tests given to students in large group testing sessions arranged two weeks prior to the end of the course. Students who were unable to attend any one of the sessions were allowed to take the test individually in the research offices.

Students participated in the project as part of the course requirement for experience in research. Each research project allowed in the course is screened by the supervisor to insure that participation offers educational benefits to the student. At the end of the testing period, each student was given a full written description of the purposes of the tests being used and the nature of the experiment, of which the current study was only one part. In this written feedback, no mention was made of the later follow-up.

TABLE 2  
RANK ORDER CORRELATIONS BETWEEN END-OF-COURSE AND FOLLOW-UP CRITERIA OF TEACHING EFFECTIVENESS

	Teaching ability ratings		Psychology courses		Attitude toward psychology		Criteria test	
	(End)	(Follow-up)	(End)	(Follow-up)	(End)	(Follow-up)	(End)	(Follow-up)
Teaching ability ratings		.94*						
	(End)							
	(Follow-up)							
Psychology courses		.14		.77				
	(End)							
	(Follow-up)							
Attitude toward psychology		.26		.94*		.66		
	(End)							
	(Follow-up)					.54	.54	
Criteria test		-.08		-.60		-.14	-.48	-.43
	(End)							
	(Follow-up)					.26	-.26	.88

N = 6 classes.

\* Significant at the .02 level (two-tail).

### *Procedures for Collecting Follow-Up Data*

Six classes were chosen for follow-up on the basis of being representative of different teaching styles but not at the extremes on any of our observational or end of course measures. Of the 152 students originally enrolled in the six classes included in our sample, we were able to locate 124 who were still in school and listed in the Student Directory 14 months after completing the course. Each of these students was sent a letter describing the research and offered \$3.00 if the student would complete the questionnaire enclosed with the letter. Of the 124 students, 92 responded. Respondents were not statistically significantly different from nonrespondents on measures completed at the end of the course, and came in about the same proportions from each of the classes in the sample. The criterion questionnaire included items chosen from:

- (1) The Student Perception of Teaching and Teacher questionnaire, evaluating the effectiveness of the teacher and course.
- (2) The Attitude Toward Psychology questionnaire.
- (3) Ten items from the Introductory Psychology Criteria Test chosen as those most discriminating between poorer and better students in previous samples. These items appeared to represent a reasonable sample of content and level of thinking required.
- (4) A questionnaire about advanced courses and other experiences related to psychology in the year since they took the introductory course.

## RESULTS

As Table 1 indicates, teachers who were effective at the end of the course in terms of student performance and ratings tended to be those most effective as measured by student responses over a year later. Since the  $N$  is small, the raw data are probably more revealing than correlation coefficients.

As Table 1 suggests, student evaluations of teaching change relatively little in the year following completion of a course, although students tend to be slightly less enthusiastic about their teachers. What is disappointing is that these ratings are not highly related to other criteria. As Table 2 indicates, the only follow-up criterion showing a substantial rank order correlation with end of course ratings was the measure of attitude toward psychology.

## DISCUSSION

We are delighted that student performance measures used as criteria of teaching effectiveness hold up very well over a 15-month retention period. The rho's of .9 and .5 are impressive even with a small  $N$  of classes.

Most research on college teaching and learning uses end-of-course measures, frequently a final examination upon which students' grades are based. We, and other students of teaching effectiveness, have been very critical of using the final examination for this purpose. In this study, the criterion measures were given just before the end of the course in a testing session not connected with final examinations and grading. Although our  $N$  was small, the strong relationship between mean scores of classes on these tests and those given 15 months later is encouraging for students of teaching effectiveness, since follow-up studies are extremely difficult and

expensive. The low correlation between student ratings and these criteria indicate that student ratings are not satisfactory substitutes for measures of student cognitive performance. They are, however, reasonably related to a measure of the achievement of attitudinal goals.

### REFERENCES

- CARRIER, N. A. *Evaluating the introductory psychology course*. Reading, MA: Addison-Wesley, 1966.
- ISAACSON, R. L., MCKEACHIE, W. J., MILHOLLAND, J. E., LIN, Y-G., HOFFELLER, M., BAERWALDT, J. D., & ZINN, K. L. Dimensions of student evaluations of teaching. *Journal of Educational Psychology*, 1964, 55, 344-351.
- MCKEACHIE, W. J., & KULIK, J. A. The effectiveness of instruction in higher education. In F. N. Kerlinger (Ed.), *Review of research in education* (Vol. 3). Itasca, IL: Peacock, 1975.
- MCKEACHIE, W. J., LIN, Y-G., DAUGHERTY, M., MOFFETT, M. M., WALZ, M., LOOMIS, M., NEIGLER, C., KERR, J., & MANN, R. *Leader responses to member hostility*. (Grant No. GS-33054 National Science Foundation.) Ann Arbor, MI: The University of Michigan, 1974.
- MILHOLLAND, J. E. Measuring cognitive abilities. In W. J. McKeachie, R. L. Isaacson, J. E. Milholland (Eds.), *Research on the characteristics of effective college teaching*. (Final report: Cooperative Research Project No. SAE 850, United States Office of Education and The University of Michigan.) Washington, DC: US Government Printing Office, 1964.