RESEARCH IN HIGHER EDUCATION
Volume 7, pages 187-189

HOW LARGE IS THE COURSE EFFECT? A Note on Romney's Course Effect vs. Teacher Effect on Students' Ratings of Teacher Competence

Gerald M. Gillmore, *University of Washington, Seattle*

Romney (1977) presented data from which he concluded that within student ratings of college instruction, the course that an instructor teaches is as important a determiner of resulting ratings as the instructor himself. Reanalysis of his data indicates that the course effect is actually quite small, a result that is consistent with earlier studies.

Key words: student ratings; course effects; teacher effects

Using a hierarchical design in which students were nested within teachers, and teachers, in turn, were nested within courses, Romney (1977) concluded that ". . . the ratings students assign to teachers in evaluating their competence is affected by the kind of course being taught and that this course effect is as strong as the teacher effect" (p. 348). Romney's conclusion deserves close attention because it is potentially very important in appropriately interpreting student ratings results especially for decisions affecting pay and promotion, and because it is contrary to three previously completed studies. Both Hogan (1973) and Bansel et al. (1975) found the correlation between sets of two courses taught by the same instructors to be much higher than the correlation between sets of two instructors teaching the same courses. Hence, they found the teacher effect to be larger than the course effect. In fact the correlations reflecting the latter tended to be quite small. Approaching the same problem from the point of view of generalizability theory,

Address reprint requests to Gerald M. Gillmore, Educational Assessment Center, University of Washington, PB-30, Seattle, WA 98105.

TABLE 1. Romney's (1977) Table 3 with Random Effects F Ratios and ω^2 Values Added

Source	D.F.	Mean Squares	F (Fixed)	P (Fixed)	F (Random)	ω^2
1st Administration:						
Course effect	7	6.4414	4.96	< 0.0001	1.27[b]	.04
Teacher effect[a]	22	5.0600	3.90	< 0.0001		.11
Error	582	1.2977				
Total	611					
2nd Administration:						
Course effect	9	3.1220	3.32	0.0005	.75[b]	.02
Teacher effect[a]	30	4.1707	4.43	< 0.0001		.11
Error	886	0.9414				
Total	925					

Note. [a] Should be written "teacher effect confounded with the teacher by course interaction"
[b] $P > .05$

Gillmore and Kane (1977) found the variance component associated with teachers to be relatively large, while that associated with courses was neglible. We concluded that "The course does not seem to be a major factor in the determination of the ratings of a class" (p 18).

Close scrutiny of Romney's results reveals that the discrepancy between his conclusion and those above do not lie in the data but in several questionable decisions and misinterpretations on his part. First, because of the nested design, he mistakenly implied that the interaction term could not be evaluated, whereas the teacher effect could be. In fact, under this design the interaction is confounded with the teacher effect and, hence, his teacher effect is, in reality, the teacher effect *plus* the teacher by course interaction.

Second, Romney cited Winer (1962, pp 184–188) in support of his method for analysis of the data. Curiously, these pages in Winer, while describing proper analyses of hierarchical designs, cover only situations in which cell sizes are equal. Romney's data include wildly unequal cell sizes, ranging from two to seventy-two students per section and from two to six teachers (sections) per course. Unfortunately, no clue is given as to the specific method of analysis used.

Third, Romney chose to treat teachers as fixed, and in making this decision was able to use the students-within-teachers mean square as his error term for testing the course main effect. In choosing this model, however, one is restricted to generalizing only to those particular teachers in the sample. Such a restriction severely limits the value

of the study. A much more interesting choice would have been to treat teachers as random, thereby allowing generalization to some larger set of teachers. If this latter model is accepted, the error term for testing the significance of the course main-effect becomes the mean square for teachers (confounded with the teacher-by-course interaction); that is, the differences among teachers within courses must be considered as an additional source of random error in assessing the course main effect. Under this model, the course effect fails to reach significance in both of his administrations. Romney's Table 3 is extended to include these random effects F ratios in Table 1.

Finally, and perhaps most importantly, Romney based his conclusion of rough equivalence between the teacher and course effect on the magnitudes of the mean squares associated with the two factors. However, mean squares are not directly indicative of the magnitude of effects but are related to significance testing. The magnitude of effects can be assessed in various ways, one of which is to compute a strength of association index, e.g., Hays' ω^2 (1963, p 381–2). Table 1 presents further extension of Romney's Table 3 adding estimates of ω^2 calculated from his data.

As can be seen, while the mean squares for the two sources are roughly equivalent, the strength of association index is much larger for the teacher effect (confounded with the teacher-by-course interaction) than for the course effect. This result is perfectly consistent with the studies mentioned above. While there may be inequities associated with student ratings of instruction, the specific course that one teaches appears to be relatively unimportant, within the natural limits imposed by current academic practices.

REFERENCES

Bausell, R. B., Schwartz, S., and Purohit, A. (1975). An examination of the conditions under which various student rating parameters replicate across time. *Journal of Educational Measurement, 12:* 273–280.

Gillmore, G. M., and Kane, M. T. The teacher and the course as units of analysis in the generalizability of student ratings of instruction. National Council of Measurement in Education, March, 1977.

Hays, W. L. (1963). *Statistics for Psychologists*. New York: Holt, Rinehart and Winston.

Hogan, T. P. (1973). Similarity of student ratings across instructors, courses, and time. *Research in Higher Education, 1:* 149–154.

Romney, D. (1977). Course effects vs teacher effect on students' ratings of teacher competence. *Research in Higher Education, 5:* 345–350.

Winer, B. J. (1962). *Statistical Principles in Experimental Design*. New York: McGraw-Hill.