RESEARCH PRODUCTIVITY AND SCHOLARLY ACCOMPLISHMENT OF COLLEGE TEACHERS AS RELATED TO THEIR INSTRUCTIONAL EFFECTIVENESS: A Review and Exploration

Kenneth A. Feldman

This analysis reviews the research that has been done on the connection between research productivity or scholarly accomplishment of faculty members and their teaching effectiveness (as assessed by their students). On average, there is a very small positive association between the two variables. To understand this relationship better, extant research was explored for factors that might mediate either positive or inverse associations between research productivity and teaching effectiveness and those that possibly could be common causes of them. Pedagogical practices and dispositions of faculty members, as well as certain course or class characteristics (size of class, electivity of course), were examined as potential mediating factors. Potential common causes investigated were academic rank and age of faculty members, their general ability, their personality characteristics, and the amount of time or effort they spend on research activities. The association between research productivity and teaching effectiveness was explored further by considering whether its size and direction varies by career stage of faculty members, their academic discipline, and the type of college or university in which they teach.

In the present series of analyses of the ratings given by college students to their teachers and courses, the search has been for those characteristics of students, of courses, and of rating circumstances that are consistently associated with these evaluations of teachers and courses (Feldman, 1976a, 1977, 1978, 1979, 1984). A corresponding interest has been in determining which characteristics, behaviors, and activities of the teachers themselves are reliably and routinely related to their perceived effectiveness in the classroom. Factors considered in this complementary area of concern have been the teacher's classroom attitudes, behaviors, and techniques along various

Kenneth A. Feldman, Department of Sociology, State University of New York, Stony Brook, NY 11794-4356.

specific pedagogical dimensions, insofar as these are perceived by students (Feldman, 1976b); the teacher's age, rank, and instructional experience (Feldman, 1983); and the teacher's personality characteristics and related traits (Feldman, 1986). The present analysis continues the focus on teachers themselves. It concentrates on how the efforts and achievements by teachers outside of the classroom – in particular, their research and scholarly productivity and accomplishments – relate to perceived effectiveness in the classroom.

As with the other analyses in the series, data for the present analysis have been restricted to studies of students and teachers at colleges and universities in the United States and Canada. Almost all of the studies included have measured the perceived effectiveness of students' present teachers at the time of data collection, by using evaluation questionnaires given to students in intact current classes. Also included in the present analysis are the few studies in which students' past teachers as well as their present ones were nominated, in some way, as superior, either by students alone or by students in conjunction with others. A sample of students, for instance, may have been asked to name the best instructors they had had at their school (rather than to evaluate their current classroom instructors), which permitted comparison of these purportedly superior instructors to others not so regarded. Research productivity and scholarly accomplishment of faculty members have been measured in several ways in the studies reviewed, including publication counts, citation counts, success in getting research funds (research support), and ratings by outside observers.

INITIAL FINDINGS: ZERO-ORDER CORRELATIONS

The present analysis begins by considering the research that compares teachers' research and scholarly productivity with students' perceptions of these teachers' *overall* instructional effectiveness. A number of reasons have been given by analysts for expecting that the research productivity and scholarly accomplishments of teachers would be positively related to their overall instructional effectiveness. At the same time, there are those who argue that these variables should be expected to be inversely associated, while still others see no reason for them to be associated at all. Discussion of these various expectations and the rationales for them is an important part of the present analysis, but will be postponed until the actual empirical relationships (if any) between research or scholarly productivity of teachers and their perceived instructional effectiveness in the classroom have been surveyed.

Table 1 presents pertinent information from existing research in the area.[1]

In most of the studies summarized in this table, at least one covariation statistic (such as a product-moment correlation coefficient) is given. Some of the studies, however, do not give specific data of this sort, although certain information about the relationship is presented. In any event, it can be seen from Table 1 that more often than not the relationship between research or scholarly productivity and overall evaluation of the teacher is not statistically significant; but note further that whether or not the relationship is statistically significant, with very few exceptions the *direction* of the relationship is *positive.*

Statistically, such "vote counting" of results is not a very powerful analysis. It may fail to identify an overall or general association between variables, especially when sample sizes are not large and associations between the variables are small or moderate in size (cf. Rosenthal, 1978, 1984, chap. 5; Gage, 1978, chap. 1; Light and Pillemer, 1984, chap. 3; and Pillemer and Light, 1980). This problem can be circumvented by averaging together the associations found in studies under consideration (in essence, determining a combined estimate of the covariation between the variables) while combining each study's individual significance test into an overall pooled test. This meta-analytic procedure of testing the significance of the combined results thus gains statistical power by using a cluster of studies bearing on the same relationship and is particularly useful when the various relationships are modest (if that) but in a consistent direction (cf. Light and Pillemer, 1984, p. 77), as they indeed are in Table 1.

The large majority of studies in Table 1 gives results in terms of a Pearson product-moment correlation coefficient (r). In those cases where associations in studies were not given in terms of r's, the associations that were presented in the studies have been converted into r's using formulas found in Glass, McGaw and Smith (1981, Table 5.8, pp. 149–150). In some studies listed in Table 1, only one correlation pertained; it is considered the summary statistic for the study. In other studies, more than one r was available for inclusion in Table 1, because there was more than one indication of overall teacher effectiveness, more than one indicator of research productivity, more than one subset of students (not combined into one overall sample), or the like. In each of these cases, an average r for the study was calculated and is taken to be the summary statistic for the study. These summary statistics, whether an individual r for some studies or an average r for other studies, are marked with asterisks in Table 1. Also marked by an asterisk are the Z's (standard normative deviates) associated with the summary r's, with p level given in parentheses.[2] All told, Table 1 contains 29 studies with summary r's and Z's. *The average correlation across these studies is* $+.12$. Although this correlation is not large, the combined Z^3 is equal to $+12.089$ and is statistically significant ($p < .001$).[4]

TABLE 1. Summary of Results of Studies Relating Research Productivity and Scholarly Accomplishments to Overall Effectiveness of Teachers as Perceived by Students

Ahern (1969): 75 teachers who received "outstanding teachers" awards from 14 New England institutions of higher education in the 5-year period beginning with the academic year 1963–1964 compared with a national profile of teachers, Spring, 1963.

The author writes that "comparing all award recipients to teachers in higher education, recipients on the average are significantly greater publishers of articles" (p. 306); and the first group also has published more books than the second (see p. 307). Information presented is insufficient for calculating an r or its equivalent.

Aiken (1975): Faculty at Guilford College, 1973 and 1974.

"Little or no relationship . . . found between course rating and scholarly productivity of the instructor" (p. 253). No further information is given.

(*) Aleamoni and Yimer (1973).

Sample 1: 360 faculty members (nominated by at least one other faculty member as deserving mention for good teaching) at the University of Illinois (Urbana-Champaign campus), 1969–1970.

Number of publications (1966–1969) × score on Overall Instructor subsection of 34-item questionnaire given to students ("The Advisor"): from Table 3, $r = -.02$ (publications unweighted by type of publication); $r = +.02$ (publications weighted by type of publication, weighting scheme I); $r = .00$ (publications weighted by type of publication, weighting scheme II). Average $r = .00$; $Z = 0.000$.

Sample 2: 28 faculty members (nominated by at least one other faculty member as deserving mention for good teaching) at the University of Illinois (Urbana-Champaign campus), 1969–1970.

Number of publications (1966–1969) × total score on 50-item Illinois Course Evaluation Questionnaire given to students: from Table 3, $r = -.04$ (publications unweighted by type of publication); $r = +.10$ (publications weighted by type of publication, weighting scheme I); $r = +.04$ (publications weighted by type of publication, weighting scheme II). Average $r = +.03$; $Z = +0.153$.

Average r (across the two average r's of the two samples weighting for sample size) = $+.002$*; average Z = (across the two Z's of the two samples weighting for sample size) = $+0.011$*.

Note. For additional information based on graduating seniors only, see Aleamoni and Yimer (1974). These data have not been included here.

(*) Bausell and Magoon (1972); 105 instructors at the University of Delaware, 1969–1970.

Number of articles for which teacher was first author × overall instructor evaluation made by students: from Table 35, $r = +.12$; for number of articles for which teacher was second author, $r < .10$ (counted here as $r = +.00$); and

TABLE 1. ***(Continued)***

for number of other publications, $r < .10$ (counted here as $r = +.00$). Average $r = +.04$; $Z = +0.406$.

Number of grants × overall instructor evaluation made by students: from Table 35, $r = +.10$; $Z = +1.019$.

Average $r = +.07$*; $Z = +0.711$*.

(*) Braunstein and Benston (1973): 349 faculty members at the University of Rochester Colleges of Arts and Sciences, Engineering, and Management, 1968–1969.

Chairman's (or Dean's) rating of faculty member in terms of research performance × students' overall evaluation of faculty member in terms of teaching effectiveness: from Table 3, $r = +.24$ for teachers in Humanities; $r = -.05$ for teachers in Social Science; $r = -.04$ for teachers in Natural Science; $r = +.36$ for teachers in Engineering; $r = -.31$ for teachers in Management.

Average $r = +.04$* (unweighted average because number of teachers in each of the five academic areas not given in the article); $Z = +0.743$*.

(*) Bresler (1968): 106 full-time faculty members at Tufts University, Fall, 1965.

Comparison of faculty members having research support with those who do not with respect to their ranking by students in the first, second, third, or fourth quartile of "teaching excellence in comparison with other Tufts faculty members" (coded with numerical designations of 1, 2, 3, and 4). Because average evaluations were given separately for junior and senior faculty within the divisions of Science and Engineering, Social Sciences, and Arts and Humanities (in Tables 1 and 2), for purposes of the present analysis average evaluations across these categories (weighting for number of faculty within a category) were determined and a t (not given in the original) was calculated: $t = +1.8515$, which converts to $r = +.23$*; $Z = +1.835$*.

Note. Standard deviations used to calculate t were taken from Tables 1 and 2, as averaged across appropriate categories (weighted for number of faculty within each category). Although nothing is said explicitly in the article, evidence internal to the tables suggests that these standard deviations are based on variation among individual students' evaluations rather than among mean evaluations of faculty members. If so, both the t and r that have been calculated are underestimates.

(*) Centra (1983, Study 1): 2,968 faculty members at 61 four-year colleges and universities, 1979. (Included were liberal arts colleges, state colleges, and a few doctoral-granting universities, most of which did not put a heavy emphasis on research.)

Number of publications in the previous 5 years × overall rating by students of teacher's effectiveness: from Table 1, average $r = +.10$* (across 3 categories of subject field of the course and 5 categories of years of teaching experience, weighted by the number of teachers in each of the 15 categories); $Z = +5.460$*.

(*) Centra (1983, Study 2): 1,623 faculty members at 10 four-year colleges and uni-

TABLE 1. ***(Continued)***

versities, 1980. (Although research universities were absent from this sample, an effort was made to include institutions with more emphasis on research productivity than those held by the institutions in Study 1; thus, only the more selective liberal arts colleges were included, along with doctoral-granting universities and a large state college.)

Number of publications in the past 5 years × overall rating by students of teacher's effectiveness: from Table 2, average $r = +.07^*$ (across 4 categories of subject field of the course and 3 categories of years of teaching experience, weighted by the number of teachers in each of the 12 categories); $Z = +2.822^*$.

(*) Clark (1973): 45 full-time faculty members at "Midwest College," a small college enrolling about 1,200 students in the arts and sciences, business, and education, Spring, 1968.

Number of publications in the prior 5 years × overall rating by students of faculty member's teaching effectiveness: $r = +.30^*$; $Z = +2.011^*$.

Cornwell (1974): 101 different lecture sections representing 70 different lecturers and 20 different college and university chemistry departments throughout the country, 1972.

"Research activity" (i.e., whether or not the lecturer actively engaged in research) × overall students' rating of teacher's effectiveness: although a point-biserial correlation can be calculated from the data presented (+.17; $Z = +1.719$ based on F), direction of the association cannot be determined.

(*) Dent and Lewis (1976): 90 faculty members from the Departments of Sociology/Anthropology, Economics, Political Science and Psychology (name of school not given), 1973.

Total number of publications × student evaluations of the teaching skill of the instructor: from Table 1, $r = +.05$; for number of citations (in the *Social Sciences Citation Index*) by colleagues within the instructor's own discipline, $r = -.01$; for number of citations by scholars outside the instructor's discipline, $r = +.02$.

Average $r = +.02^*$; $Z = +0.188^*$.

(*) Faia (1976): 53,034 full-time faculty members at 301 institutions of higher education, 1972–1973.

For institutions weak on research emphasis (comprehensive universities and colleges not offering the doctorate, liberal arts colleges, and two-year colleges and universities), when 5 × 2 contingency table (see Table II) is collapsed into 2 × 2 contingency table (yes/no publications within prior two years × yes/no ever received teaching award), tetrachoric $r = +.15$ ($N = 265{,}682$ when each respondent was weighted so that the sample would be representative of the entire national population of college and university faculty members).

For institutions strong on research emphasis (research universities and doctoral-granting universities), when 5 × 2 contingency table (see Table II) is col-

TABLE 1. ***(Continued)***

lapsed into 2 × 2 contingency table (same as above), tetrachoric $r = +.07$ ($N = 206{,}979$, when weighted as explained above).

Average $r = +.11^*$ (across the r's of the two samples, whether or not weighting by the weighted N's); $Z = +25.408^*$ (based on $N = 53{,}034$).

(*) Freedman, Stumpf, and Aguanno (1979); also Stumpf, Freedman, and Aguanno (1979): 129 instructors teaching 334 classes in "a large northeastern university," year(s) not given.

Number of publications over three years (adjusted for the annual rate if less than three years) weighted by type of publication × overall rating by students of instructor (adjusted for grading policy of instructor): from Table 1, $r = +.23^*$; $Z = +2.626^*$.

(*) Frey (1978): 36 professors and 6 associate professors in biochemistry, biology, chemistry, geology, and physics who had taught at least one undergraduate class in the Fall or Winter quarter of 1975–1976.

Number of citations × student evaluation of teacher on skill factor scale: $r = +.37$ (see p. 83); for student evaluation of teacher on rapport factor scale, $r = -.23$ (see p. 83).

Average r (across the two r's for the two factor scales) $= +.07^*$; $Z = +0.443^*$.

(*) Friedrich and Michalak (1983); Michalak and Friedrich (1981): Faculty members at Franklin and Marshall College, a small liberal arts college in Pennsylvania, 1977–1978.

Evaluation by dean of the college and the chairperson of the faculty member's department of research of 74 faculty members (based on the amount and quality of publications, research in progress, systematic programs of study and involvement in professional activities) as averaged over a five-year period (1972–1977) × students' rating of the faculty member's overall teaching effectiveness: $r = +.17$ (see Friedrich and Michalak, 1983, p. 152); $Z = +1.453$; for number of citations in the *Science Citation Index* and the *Social Science Citation Index* over a five-year (1972–1977) × evaluation by the dean of the college and the chairperson of the faculty member's department of teaching merit of 53 faculty members in the disciplines of natural sciences and social sciences (based on student evaluations of courses the faculty member teaches, exit interviews with departing seniors, "grapevine" feedback from students, examination of course syllabi, and in some cases firsthand observation of the person's teaching), $r = +.20$ (see Michalak and Friedrich, 1981, p. 590); $Z = +1.443$.

Average r (across the two r's weighting for number of faculty members) = $+.18^*$; average Z (across the two Z's weighting for number of faculty members) $= +1.499^*$.

Note. The correlation of $+.33$ given on p. 589 in Michalak and Friedrich (1981) has not been averaged in because it is based on research merit ratings

TABLE 1. ***(Continued)***

and teacher merit ratings made by the same raters (the chairperson of the faculty member and the dean of the college) and is likely to be artificially large.

Goldsmid, Gruber, and Wilson (1977): 39 nominees for distinguished teaching awards compared with all colleagues in the same department of the same academic rank at the University of North Carolina–Chapel Hill, 1972–1973 and 1973–1974.

The authors write that "while the findings were not statistically significant ($p = .10$), for most of the nominees, research publication is about the same as, or better than, that of their counterparts" (p. 434); r or its equivalent cannot be determined from information given.

Grant (1971): 685 faculty members at the University of Utah, 1968.

As the faculty members' percentage of time allocated to nonsponsored (nonfunded) research and writing increased, the courses they taught received lower mean scale scores on "recommended" item of an evaluation form completed by students. For sponsored (funded) research, no statistically significant relationship was found. In both cases, r or its equivalent cannot be determined from information given (on p. 406).

(*) Harry and Goldner (1972): 211 full-time faculty members of the Colleges of Liberal Arts and Education at "a large public urban midwestern university," 1968–1969.

Number of instructor's published articles plus three times number of published books × percentage of students in the class giving the instructor an "A" (overall evaluation of his or her performance): from Table 2, $r = +.19^*$; $Z = +2.772^*$.

Hayes (1971)[a]*:* 355 faculty members in 17 of the academic departments at Carnegie-Mellon University, Fall, 1971–Fall, 1968.

Number of publications for the prior five years (weighted by type of publication and adjusted for faculty members for whom fewer than five productive years had elapsed since they had received their Ph.D.) and grant status (whether faculty member had a grant during Fall, 1967–Spring, 1969) × either how stimulating the class was (Fall, 1967 and Spring, 1968) or how stimulating the teacher's presentation of the course material was (Fall, 1968). These specific evaluations are not considered an overall evaluation of the teacher for the present analysis. Results are coded in the Appendix (see Instructional Dimension No. 1).

(*) Hicks (1974): 459 faculty members at San Jose State College, year(s) not given.

Teachers who had published compared with teachers who had not on student ratings: $t = +4.19$ given in Table 1 converted to a point-biserial correlation of $+.19$ and an r of $+.25^*$; $Z = +4.150^*$ (based on $t = +4.19$, $df = 457$).

(*) Hoffman (1984b): 65 faculty members in the College of Education at Florida Atlantic University, 1980.

TABLE 1. ***(Continued)***

Research productivity for the calendar year (publications, works in special form as musical and artistic products, grants, attendance at and participation in professional meetings, prizes and awards, as numerically weighted consistent across departments) × student evaluations of teacher (single mean score for all courses evaluated during the calendar year): from Table 1, $r = -.25^*$; $Z = -2.015^*$.

Note. For an analysis of data from a subsample ($N = 16$) of the most and least productive in research of the 65 faculty members, see Hoffman (1984a).

Hoyt (1974)[a]*:* 173 faculty members at Kansas State University (most of the teachers instructing undergraduate course in the Colleges of Education, Commerce, Architecture and Design, Agriculture, and Home Economics plus about half of those teaching courses in Arts and Sciences).

Number of publications (weighted by type of publication and divided by the number of years the teacher had been employed full-time at the university) × students' perceived progress on eight instructional objectives. This measure is not considered an overall evaluation of the teacher for the present analysis. Results are coded in the Appendix (see Instructional Dimension No. 12).

(*) Hoyt and Spangler (1976): 183 faculty members at Kansas State University who voluntarily participated in that institution's student evaluation of instruction program and who were in the natural-mathematical sciences or social-behavioral sciences), 1972–1973.

Research involvement (time commitment and accomplishment) as rated by the department head of the faculty member (considerable, moderate, and little) × overall liking of instructor (degree to which student would like to take another course with the instructor controlled for student's motivation to take the course): from Table II, r (eta) $= +.17^*$ (F of 2.642 converted to eta, see Glass, McGaw, and Smith, 1981, Table 5.8); $Z = +1.619^*$ (based on F, see Rosenthal, 1984, p. 107 and Formula 5.15).

Lasher and Vogt (1974): Faculty members at the College of Business Administration, Bowling Green State University, 1969.

Involvement in nonteaching activities (research and/or public service) × overall evaluation of teacher (letter grade assigned to the teacher by the student). Value of chi-square is statistically significant, but neither r nor the direction of the association can be determined (see Exhibit I).

(*) Linsky and Straus (1975): Faculty members at 16 colleges and universities, year(s) not given.

Number of publications (for 1,422 faculty members) over an approximately 20-year period weighted by type of publication × overall rating of professor's classroom teaching performance (obtained by researchers by combining the average ratings on several rating items): from Table 1, $r = +.04$; $Z = +1.508$; for number of citations (for 760 faculty members) in the *Science Citation Index* over a 10-year period, $r = -.05$, $Z = -1.377$.

TABLE 1. ***(Continued)***

Average r (across the two r's weighting for number of faculty members) = +.01*; average Z (across the two Z's weighting for number of faculty members) = +0.503.*

(*) *Marquardt, McGann, and Jakubauskas (1975):* 91 course sections in the College of Commerce and Industry, Fall, 1972.

Number of academic publications appearing in publications not affiliated with the institution × mean scale value (for the class section) of students' responses to ten 5-point scales measuring components of the teacher's performance: from Table 1, $r = +.25$*; $Z = +2.395$*.

(*) *Marsh and Overall (1978):* 183 faculty members in the Division of Social Sciences at the University of Southern California, 1977–1978.

Faculty member's self-rating on 5-point scale of the "scholarly production in their discipline" × overall rating by students of the teacher: from Appendix VI, $r = +.14$*; $Z = +1.893$*.

(*) *McCullagh and Roy (1975):* 52 faculty members at Appalachian State University, 1971–1972.

Number of academic articles published during a one-year period × students' rating of general classroom teacher effectiveness: from Table 2, $r = +.03$; for number of academic books published during a one-year period, $r = +.06$.

Average $r = +.05$*; $Z = +0.354$*.

(*) *McDaniel and Feldhusen (1970):* 76 faculty members at Purdue University, year(s) not given.

Number of books for which faculty member was first author × composite instructor evaluation by students: from Table 1, $r = -.13$; for number of books for which faculty member was not first author, $r = +.24$; for number of articles for which faculty member was first author, $r = -.10$; for number of articles for which faculty member was not first author, $r = +.16$.

Average $r = +.04$*; $Z = +.344$*.

Plant and Sawrey (1970): 32 faculty members "in the psychology department of a large tax-supported state college in California," year(s) not given.

Number of scholarly publications of faculty member, number of papers presented at meetings of scientific-professional societies, whether faculty member had received financial support for research activities from any source for a prior three-year period, whether faculty member had received financial support for research activities from any source outside the university during the three-year period × students' evaluation of faculty member on eight specific rating items and the total score across these items. The investigators report that very few of the many associations generated were statistically significant, but exact data are not given.

Ratz (1975): 15 faculty members in Electrical Engineering at the University of Waterloo (Canada), 1969–1975.

TABLE 1. ***(Continued)***

Number of awards from the National Research Council of Canada during 1972–1975 × average student evaluation across a ten-item questionnaire over the 1969–1973 period. Author writes that a "two-by-two contingency table obtained from the medians in each quantity failed to reveal any association whatever, between student ratings and research grants" (p. 126). Information given is insufficient for calculating an r or its equivalent.

Riley, Ryan, and Lifshitz (1950)[a]*:* 382 teachers at Brooklyn College, Spring, 1947.
Whether or not the teacher has published × students' evaluation on specific evaluation items. Although there was an overall evaluation item on the questionnaire given to students, the results for this item are not reported. Results for specific evaluations of teachers are given in the Appendix (see Instructional Dimensions No. 4, No. 5, No. 6, and No. 18).

Rossman (1976)[a]*:* 122 full-time faculty members at Macalester College (a private liberal arts college in St. Paul, Minnesota), 1969–1970.
Rating of each faculty member by three members of the faculty personnel committee on publication productivity (based on public information from faculty member's personnel folders) × number of times faculty member was nominated by students as teaching the most stimulating course they had had and as that teacher who had contributed the most to their education and/or personal development. These specific evaluations are not considered an overall evaluation of the teacher for the present analysis. Results are coded in the Appendix (see Instructional Dimensions No. 1 and No. 12).

(*) Rushton, Murray, and Paunonen (1983, Study 1): 52 full-time psychology professors of varying ranks who were, or recently had been, at the University of Western Ontario, 1974–1979.
Number of publications produced over a four-year period, 1976–1979, as listed in the *Source Index* of either the *Social Science Citation Index* or the *Science Citation Index* × students' rating of the "overall effectiveness" of the instructor: from p. 103, $r = +.10$; for number of citations (excluding first authored self-citations) over a four-year period, 1976–1979, as listed in the *Social Science Citation Index,* $r = -.24$.
Average $r = -.07$*; $Z = -0.496$*.

(*) Siegfried and White (1973): 45 faculty members in the Economics Department at the University of Wisconsin–Madison, 1971–1972.
Total number of monographs (all published books excluding textbooks and edited volumes) × overall evaluation of the teacher (whether student would recommend the professor to a friend) as adjusted for course level and seasonal variation: from Table 2 in Appendix II, $r = +.02$; for total number of articles in national, general journals, $r = +.02$; for total number of articles in specialty and regional journals, $r = +.13$; for total number of all other publications (including textbooks, edited books, chapters in books, research reports, etc.), $r = -.05$.
Average $r = +.03$*; $Z = +0.197$*.

TABLE 1. ***(Continued)***

(*) *Stallings and Singhal (1970, Study 1):* 128 faculty members at the University of Illinois, 1965–1966.

Total number of publications weighted by type of publication × total score on the Course Evaluation Questionnaire completed by students: from p. 142, $r = +.26^*$; $Z = +2.968^*$.

(*) *Stallings and Singhal (1970, Study 2):* 121 faculty members at Indiana University, 1967.

Total number of publications weighted by type of publication × overall comparative ranking of the instructor by students: from p. 142, $r = +.13$; for global index of teaching effectiveness (average of eight items), $r = +.08$.

Average $r = +.11^*$; $Z = +1.204^*$.

(*) *Stavridis (1972):* 32 faculty members primarily involved in teaching undergraduates in the College of Education at Arizona State University.

Total number of publications weighted by type of publication × overall student rating of the instructor in general (all-around) teaching ability: from Table 9, $r = +.27$; for overall student rating of the instructor compared with other instructors at the university, $r = +.20$.

Average $r = +.24^*$; $Z = +1.333^*$.

Teague (1981): 9 outstanding teaching award recipients compared with 9 nonrecipients at the University of Maryland, College Park Campus (two-year period, exact years not given).

The outstanding teaching award recipients on the average were slightly lower on the number of books published, presentations given, and grants received, and slightly higher on number of articles published; but none of the differences were statistically significant. Data presented (in Table 1) are insufficient for calculating an r or its equivalent.

(*) *Usher (1966):* 26 full-time faculty members in the College of Education at the University of Florida, 1965.

Dean's rating of faculty members research and publications (based on the two most recent annual reports submitted to the dean) × overall student rating of instructor in general (all-around) teaching ability: from Table 10, $r = +.23^*$; $Z = +1.142^*$.

Voeks (1962): 193 teachers from 28 departments at the University of Washington, 1948–1952.

Publications for a five-year period (1948–1952) and for teacher's whole professional life (in both cases weighted by type of publication) × overall evaluation of teacher. In both cases, associations are not statistically significant. Data provided in Table 2 are insufficient for calculating an r or its equivalent, and the direction of the association cannot be determined.

(*) *Wood and DeLorme (1976):* 69 faculty members in the College of Business Administration at the University of Georgia, 1971–1973.

TABLE 1. ***(Continued)***

Number of publications during the period 1971–1973 × student evaluations of the instructor's effectiveness: $r = +.39$*; $Z = +3.354$* (as calculated from the t of 3.516 given on p. 78, see Rosenthal, 1984, p. 187, formula 5.15).

(*) *Wood (1978):* 23 faculty members of the Department of Educational Foundations and Inquiry in the Department of Education at Bowling Green State University, 1974–1977.

Weighted combination of books, articles and papers, research reports, book reviews, and grants of previous three-year average student evaluation of instructor based on the three most general questions of the *Student Description* form: from Table 6, $r = -.07$*; $Z = -.321$*.

Note. An asterisk in parentheses before the citation to a study indicates that the study has data that could be used in the meta-analytic procedures described in the text proper. For each of these studies, the effect size—either a single r, or an average r if more than one correlation was pertinent for a study—is marked with an asterisk. These asteriskized effect sizes are the component r's that have themselves been averaged in the meta-analyses reported in the text. When the associations in any of these studies were not given in the form of a product-moment correlation, the statistics that were given were converted to r's using the procedures suggested by Glass, McGaw, and Smith (1981, see especially Table 5.8, pp. 149–150). The Z (standard normal deviate) associated with each component r is also marked with an asterisk. These are the component Z's, from which combined Z's were calculated, and were obtained by using procedures suggested in Rosenthal (1984, see especially pp. 106–107).

[a]Four studies—Hayes (1971), Hoyt (1974), Riley, Ryan, and Lifshitz (1950), and Rossman (1976)—whose populations are described in this table have data on research productivity but not on *overall* effectiveness of teachers as perceived by students (although one or another of them is sometimes said to). As the entry for each of them explains, the data in it pertain to the *specific* evaluations of teachers and are therefore reported in the Appendix.

Because studies vary in their indicators of research and scholarly productivity, it is of interest to see whether the association between productivity and perceived effectiveness varies by type of indicator.[5] Of the 29 studies with summary r's, 21 use the number of scholarly publications of the faculty member to measure productivity. The average correlation coefficient across these studies is $+.13$ (combined $Z = +13.132$; $p < .001$). Of these 21 studies, 10 measure the teacher's *current productivity* (of the more immediate past, the exact number of years varying by study),[6] while 11 of them focus on "life-time" or *total career productivity.*[7] Surprisingly, perhaps, these two subsets of studies, on average, show the same results. The average correlation across the first set of studies is $+.13$ (combined $Z = +13.418$; $p < .001$), and that for the second set is $+.14$ (combined $Z = +5.351$; $p < .001$).

For two studies measuring research support—either number of grants received (Bausell and Magoon, 1972) or whether the faculty member had received a grant (Bresler, 1968)—as an indicator of research productivity, the average correlation is $+.17$ (combined $Z = +2.017$; $p = 0.44$).[8] Across four other pieces of research (Braunstein and Benston, 1973; Friedrich and Michalak, 1983; Hoyt and Spangler, 1976; and Usher, 1966), average correlation was also positive, and to about the same degree ($r = +.15$; combined $Z = +2.479$; $p = .013$), between the rating of research productivity of a faculty member made by the faculty member's department head (either alone or in conjunction with the dean of the college) and perceived teaching effectiveness of the faculty member. Five studies (Dent and Lewis, 1976; Frey, 1978; Linsky and Straus, 1975; Michalak and Friedrich, 1981 [in Table 1, see entry for Friedrich and Michalak, 1983]; and Rushton, Murray, and Paunonen, 1983) used citation counts of faculty members' published material, which is considered an indicator of the *quality* of research productivity in contrast to its *quantity*. This aspect of research productivity appears *not* to be related to perceived teacher effectiveness, for the average r is $-.002$ (combined $Z = -0.500$; $p = .617$).[9]

In all, then, research productivity is positively but very weakly correlated with overall teaching effectiveness (as assessed by students). It seems to make little difference what indicator of productivity is used, with one exception: The two variables appear to be unrelated when citation counts are used to measure productivity. Frey (1978) has suggested one possible reason for either weak associations or the lack of associations. His reasoning is based upon a consideration of *specific* evaluations of teachers as contrasted with overall or global evaluations, and the possibility of differential and contrasting relationships of these specific evaluations with productivity. From a larger study of teacher evaluations at Northwestern University, he analyzed data for a smaller subset of 36 full professors and 6 associate professors of chemistry, biology, chemistry, geology, and physics. Finding that the number of citations to the work of these teachers was correlated positively with a multi-item factor scale measuring the "pedagogical skill" of the instructor ($r = +.37$) but negatively with a multi-item factor scale measuring the instructor's "rapport" with students ($r = -.23$), he concluded: "it is common for instructional rating instruments to include both skill and rapport items. The net effect of combining one set of items which correlates positively with research productivity with a second set which correlates negatively is to mask the true relationships" (p. 83). This same reasoning would also apply to single overall rating items (as well as total scores on multi-item scales), since a student's overall evaluation is presumably based on the student's "aggregating" across his or her impressions in a variety of specific instructional areas, including skill and rapport.

In order to see whether additional evidence could be found to substantiate Frey's conclusion, studies comparing research productivity with specific evaluations of teachers—either in addition to or instead of overall evaluation—were drawn together for analysis. Summaries and results of these studies are given in the Appendix, where the various specific pedagogical attitudes and practices of teachers as well as other specific instructional aspects of the course—as assessed by students—have been clustered together into specific instructional dimensions and numbered as Instructional Dimensions No. 1 through No. 19. These dimensions are nearly identical to the ones that were developed in a much earlier analysis of the particular practices and characteristics of teachers that are felt by students to differentiate superior teachers and teaching from less effective teachers and teaching (Feldman, 1976b); and they are identical to the dimensions recently found to be useful in the analysis of the seniority and experience of college teachers and the size of the courses taught, as these factors relate to the specific evaluations of teachers (Feldman, 1983, 1984).

Table 2 summarizes the results given in the Appendix. For each dimension, the associations in the various studies have been averaged, and the individual significance tests have been combined into an overall pooled test. Because some of the rating items or scales in some of the studies encompass more than one of the instructional dimensions, the associations between research productivity and these items or scales appear in more than one dimension in the Appendix and thus are used more than once in arriving at averages. So as not to "overcount" such associations when averaging, they are weighted by the inverse of the number of dimensions for which they are relevant (i.e., a weight of ½ if they were coded into two of the dimensions, ⅓ if three dimensions, etc.). These weights are given in the Appendix.

To take an example of this weighting procedure, consider the "Instructor Enthusiasm" Scale in the study by Marsh and Overall (1979). This scale has specific evaluation items pertaining to the instructor's being able to make the course material interesting as well as items about the instructor's enthusiasm. Thus the correlation of +.02 between research productivity and evaluation on this scale is relevant for, and is coded in, both Instructional Dimensions No. 1 ("Teacher's Stimulation of Interest in the Course and Its Subject Matter") and No. 2 ("Teacher's Enthusiasm") in the Appendix. The average correlation between productivity and specific evaluations of teachers for each of the 19 instructional dimensions given in Table 2 takes such multiple coding into account. In this particular example, the correlation of +.02 is weighted by ½ in each of the two instructional dimensions in which it appears. If a correlation could fit into one and only one instructional dimension, its weight, obviously, is unity. Thus three correlations (+.04, +.19, +.07) of the four correlations coded in Dimension No. 2 (in the Appendix) could be coded only there, and each has a weight of one. The fourth correlation of +.02, as just noted, has a weight of ½. So the average (weighted) correlation across the correlations from the four different studies (given in the Appendix) with data relevant to Dimension No. 2 is +.09, obtained as follows:

$$[+.04 + .19 + .07 + \tfrac{1}{2}(.02)]/3\tfrac{1}{2} = +.09.$$

TABLE 2. Average Correlation between Research or Scholarly Productivity and Student-Rated Effectiveness of Teacher on Various Instructional Dimensions

Instructional Dimension	(Weighted) N	Average (Weighted) r	Combined (Weighted) Z	p
No. 1. Teacher's Stimulation of Interest in the Course and Its Subject Matter	6⅚	+.08	+2.001	.045
No. 2. Teacher's Enthusiasm (for Subject or for Teaching)	3½	+.09	+1.518	.129
No. 3. Teacher's Knowledge of the Subject	5	+.21	+6.618	<.001
No. 4. Teacher's Intellectual Expansiveness (and Intelligence)	2	+.15	+2.330	.020
No. 5. Teacher's Preparation; Organization of the Course	5⅓	+.19	+5.307	<.001
No. 6. Clarity and Understandableness	7⅙	+.11	+3.785	<.001
No. 7. Teacher's Elocutionary Skills				
No. 8. Teacher's Sensitivity to, and Concern with, Class Level and Progress	1¼	+.07	+1.166	.244
No. 9. Clarity of Course Objectives and Requirements	3½	+.18	+3.172	.002
No. 10. Nature and Value of the Course Material (Including Its Usefulness and Relevance)	4½	+.06	+2.064	.039
No. 11. Nature and Usefulness of Supplementary Materials and Teaching Aids	2	+.08	+2.416	.016
No. 12. Perceived Outcome or Impact of Instruction	5⅚	+.10	+2.524	.011
No. 13. Instructor's Fairness; Impartiality of Evaluation of Students; Quality of Examinations	4⅚	−.001	+0.773	.464
No. 14. Personality Characteristics ("Personality") of the Instructor	1	+.12	+1.022	.307

TABLE 2. *(Continued)*

No. 15. Nature, Quality, and Frequency of Feedback from the Teacher to Students	1½	+.07	+0.849	.396
No. 16. Teacher's Encouragement of Questions and Discussion, and Openness to Opinions of Others	3 7/12	−.0005	+0.617	.537
No. 17. Intellectual Challenge and Encouragement of Independent Thought (by the Teacher and the Course)	2 7/12	+.09	+3.024	.003
No. 18. Teacher's Concern and Respect for Students; Friendliness of the Teacher	4	+.05	+0.326	.744
No. 19. Teacher's Availability and Helpfulness	4 7/12	−.0004	+0.915	.360

Note. This table is constructed from data given in the Appendix. It is based on those studies in which associations between research productivity and students' specific evaluation of teachers were given either in the form of a product-moment correlation or in a form convertible into a product-moment correlation (see pp. 147–151 in Glass, McGaw, and Smith, 1981.) Because some of the items or scales measuring specific evaluations fit into more than one dimension, certain of the associations between research productivity and specific evaluations were coded into more than one of the instructional dimensions and were weighted by the inverse of the number of dimensions into which they were coded (a weight of 1 if 1 dimension, ½ if 2 dimensions, ⅓ if 3 dimensions, etc.) These weights are given in the Appendix, and the weighting procedure is more fully explained in the text proper. Adding up the weights for a dimension produces the (weighted) *N* for that dimension. The *Z*'s of the individual associations have been weighted in the same way as the *r*'s. For each dimension, the method of adding weighted *Z*'s described in Rosenthal (1984, pp. 94–95) has been used for combining probabilities—again, as more fully explained in the text—and is given in the column headed "Combined (Weighted) *Z*." The probability levels (*p*'s associated with the *Z*'s are two-tailed because the direction of each average correlation had not been predicted.

The *Z*'s, likewise, are weighted appropriately, and the method described in Rosenthal (1984, pp. 94–95) of adding weighted *Z*'s is used for combining probabilities. This method consists of multiplying each *Z* by its weight, adding the products together, and then dividing by the square root of the sum of the individual weights each squared. Thus, for Dimension No. 2, the combined $Z = [0.340 + 0.580 + 1.682 + \frac{1}{2}(0.269)]/\sqrt{1^2 + 1^2 + 1^2 + (\frac{1}{2})^2}$, which equals $+2.737/\sqrt{3\frac{1}{4}}$, or +1.518.

Table 2 gives the average weighted *r*'s, combined *Z*'s, and *N*'s (along with *p* values) for the associations between research productivity and ratings on each of the instructional dimensions. The four dimensions having the largest average correlations with research productivity are those pertaining to the teacher's knowledge of the subject matter (Dimension No. 3: $r = +.21$) and intellectual expansiveness (No. 4: $r = +.15$), the teacher's preparation and organization of the course (No. 5: $r = +.19$), and the clarity of the course objectives and requirements (No. 9: $r = +.18$). Research productivity is also positively related (with statistical significance) to six other specific instructional dimensions, although with less magnitude: the teacher's clarity and understandableness (No. 6: $r = +.11$); encouragement of independent thought as part of the intellectual challenge of the course (No. 17: $r = +.09$) and stimulation of interest (No. 1: $r = +.08$); outcome or impact of the instructor (No. 12: $r = +.10$); and the nature, value, and usefulness of the major course material (No. 10: $r = +.06$) as well as the supplementary material and teaching aids (No. 11: $r = +.08$).

The remaining dimensions of specific evaluations of teachers are unrelated to research productivity (average r's being statistically insignificant). Thus, the teacher's enthusiasm, sensitivity to (and concern with) class level or progress, and availability and helpfulness are not related to research or scholarly productivity. Neither are the teacher's encouragement of questions and discussion (and openness to others' opinions), concern and respect for (and friendliness with) students, fairness and impartiality of evaluation, and the nature, quality, and frequency of feedback to students. The overall "personality" of the teacher as perceived by students (Dimension No. 14) is also unrelated to research productivity, but this particular result is based on only one study, and not much importance should be attached to it.

Taken as a whole, these results provide little support for the hunch that different instructional dimensions are related to research productivity in sharply contrasting ways. Thus the average correlations between research or scholarly productivity and specific evaluations ranged from essentially zero to a high of $+.21$; none were inverse. Returning to the study by Frey (1978), it may be noted that the specific evaluation items loading highest on the "pedagogical skill" factor scale of his research—the scale that was positively correlated with research productivity in his study—were "advanced planning," "presentation clarity," and "increased knowledge." In the present analysis, these items refer to Dimensions No. 5, No. 6, and No. 12, each of which dimensions was found to be positively correlated with research productivity, as already shown (average $r = +.19$, $+.11$, and $+.10$, respectively). The specific evaluation items that loaded highest on the "rapport" factor scale in Frey (1978)—which scale was inversely related to research productivity in his study—were "grade accuracy," "class discussion," and "personal help." Each dimension into which these are codable in the present analysis (No.'s 13, 16, and 19) was found to be essentially unrelated to research productivity (average $r = -.001$, $-.0005$, and $-.0004$, respectively), thus differing from Frey's finding of an inverse relationship between the factor scale of "rapport" (containing these items) and research productivity.[10]

There is more to be said about the associations between research productivity and specific evaluations of teachers. This will be done, as part of the next section of this paper, in a context that is both more comprehensive and more explicitly theoretical than the one being used at present. For the moment, the point of interest is that extant research does show some variation in the extent to which (and whether) evaluations of teachers along different specific dimensions correlate with research productivity, but this variation is not nearly sharp enough or different enough in *direction* of associations to explain why the correlation between evaluation of overall teaching effectiveness and research productivity is generally so small.

Other explanations have been offered for the low association found between research productivity and teaching effectiveness. Harry and Goldner (1972) and Linsky and Straus (1975), for example, note the possibility of a curvilinear relationship between research productivity and teaching effectiveness. If the relationship in actuality is thus nonlinear, a correlation coefficient, being a measure of linear relationship, would not "capture" actuality well and would most likely be small in size, if not zero. Harry and Goldner (1972) did not find a curvilinear relationship in their research, which is as much as can be said at this point, since the other studies located for the present review did not check for curvilinearity (one way or the other). Another explanation for the small associations found hinges on the notion that teaching effectiveness probably varies widely within colleges as well as between them, and that research productivity of faculty members also varies quite a bit across colleges but much less so within each college (cf. Faia, 1976; and Hammond, Meyer, and Miller, 1969). If research productivity tends to be relatively similar among faculty members of a particular college or university (or among a particular subset of colleges or universities), then the degree to which research productivity and teaching effectiveness will show a correlation is restricted by this curtailed variance in one of the variables. Again, this possibility has not been systematically explored in extant research and could well be placed on future research agendas.

One other explanation for the typically small size of the correlation between research or scholarly productivity and teaching effectiveness is that forces creating a positive correlation between them are more or less counterbalanced by forces producing a negative correlation. It is this possibility that is examined at some length, and in a broader context, in the next section of the present analysis.

FURTHER EXPLORATIONS

On the basis of the small positive correlation (on average) between research or scholarly productivity of faculty members and their student-perceived effectiveness as teachers, it could be maintained that research productivity seems to have a very slight likelihood of benefitting teaching. But it could just as well be argued that the correlation is generally so small that, for all practical purposes, the variables in question are independent of each other. Either way, it would seem obvious that performance as a researcher or scholar outside the classroom does not significantly detract from performance as a teacher in the classroom or interfere with effective teaching (insofar as this teaching is assessed by students). Thus, Centra (1983) is not alone in concluding that "the lack of consistent negative correlations between research productivity and teacher ratings in this and other studies indicates that performance as a scholar or researcher does not significantly detract from performance as a teacher" (p. 388). For instance, Hicks (1974) also asserts that "while RP [research productivity] may not be of great benefit [to teaching], it certainly does not seem to interfere with effective teaching" (p. 145); and similar sentiments can be found in Linsky and Straus (1975), Gaff and Wilson (1971), and Thielens (1971).

Yet conclusions in this area should not be drawn too quickly. That research productivity at least does not detract from teaching and might even have a slight likelihood of benefiting it may indeed be true, but this conclusion does not necessarily follow from the observed positive (but small) correlations that have been found, for it is conceivable that relatively effective teachers who are also productive in research would be even more effective were they to do less research. As Black (1972) puts it, "superior faculty may well do research and teach better than inferior faculty, but they might teach even better if they did no or less research" (p. 349). This statement implies that research productivity essentially hinders teaching, even though there is an observed positive correlation between productivity and instructional effectiveness. At first glance, this seems puzzling, if not contradictory, but such a state of affairs is possible. One way it could come about would be the circumstance in which an essentially negative effect of research productivity on teaching evaluation was masked by other conditions or forces creating an observed positive correlation between the two.

As a somewhat oversimplified example, suppose research productivity essentially lowers teaching effectiveness by making teachers less available to students. This negative relationship could be just barely counterbalanced, or even slightly more than counterbalanced, by the fact that it is the brighter and more intelligent faculty members who are productive in their research and who are also better teachers (and seen as such by students), thereby producing either no correlation between research productivity and teaching effectiveness or a small positive correlation between them, respectively. In this example, being productive in research actually detracts from teaching, even though there is no observed relationship between the two (the first case) or the observed relationship is weakly positive (the second case). Thus, all other things being equal, if the better teachers were to do less research, they would actually be more effective in their teaching.

By the opposite token, however, just the reverse might be true. That is, research productivity by itself might have strong benefits for teaching, but these benefits may be counteracted or offset by other factors or circumstances, thus reducing, or altogether suppressing, what would otherwise be a relatively large positive correlation between research and teaching (cf. Michalak and Friedrich, 1981; and Friedrich and Michalak, 1983). To take an example, again oversimplified, research productivity may have a fairly sizable positive effect on teaching effectiveness by helping to create teachers who are more knowledgeable and more interesting in class, but this otherwise positive effect is reduced in size, or even completely offset, by some other factor, say a personality trait in individuals that facilitates productivity in research but hinders effectiveness in teaching.

These examples hardly exhaust the possibilities. As the matter thus becomes more complex, it is all the more important to actuate the logic of data analysis. That is, it is essential to search out and consider the possible availing and countervailing forces that may be at work as well as to sort out causal from noncausal relationships. In this way, factors that may mediate the causal effects of research productivity on teaching effectiveness can be distinguished from "extraneous" factors that may merely produce the *appearance* of a causal relationship between research productivity and teacher effectiveness by affecting each one separately. The first set of factors is sometimes said to "interpret" a relationship, while the second set is said to "explain" it (see, for example, Hyman, 1955; and Cole, 1976).

Figure 1 shows some of the possibilities involved in a more systematic way. In Figure 1a, research productivity is pictured as somehow *directly* affecting teaching effectiveness, while in Figure 1b research productivity is seen as influencing teaching effectiveness through its influence on one or more intervening variables (i.e., various mediating mechanisms and processes).[11] Signs on the arrows can be positive or negative, indicating either positive or inverse causal relationships. It is possible, of course, that research productivity tends to affect teaching positively through some intervening factors and inversely through others, so that the resulting correlation approaches zero as the positive and negative influences come increasingly to counterbalance each other.

In Figure 1c, research productivity and teacher effectiveness have one or

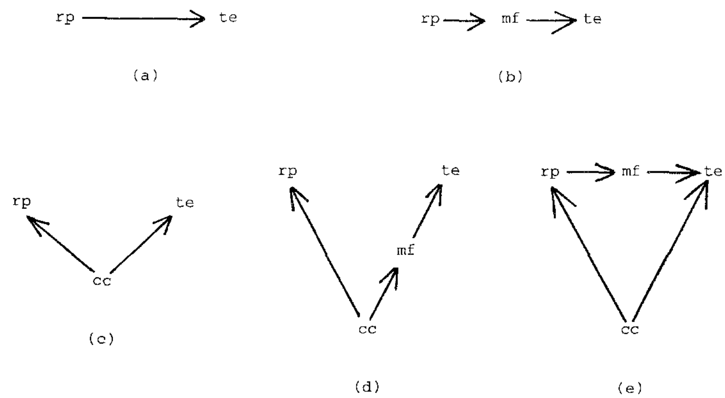

FIG. 1. Some possibilities as to how research productivity (rp), teaching effectiveness (te), and other factors are causally related; mf refers to mediating factor(s), and cc to common cause(s).

more causes in common, so that they will be correlated with each other even though neither causes the other (either directly or indirectly through intervening variables). Indeed, any conclusion of causality between research productivity and teaching effectiveness would be said to be "spurious." Again, the association of the common-cause variable with either research productivity or teaching effectiveness can be positive or negative. Moreover, the common-cause variable(s) can affect either or both of the two variables in question through one or more intervening variables (see Figure 1d).[12]

Of course, both a set of variables intervening between research productivity and teacher evaluations and a set of common-cause variables may be involved in an observed correlation between research productivity and teaching effectiveness, as shown in Figure 1e. These two sets of variables may act in concert, or they may not. An example of the latter circumstance would be when the common-cause variables produce an inverse relationship between research productivity and teaching effectiveness while research productivity is positive affecting teaching (indirectly through its effect on intervening variables). In this case, although strong influences may be at work, they may balance out in such a way that the resultant zero-order association between research productivity and teaching effectiveness is only weakly positive or negative, or there may be no association observed between the two at all.

Perhaps more than other researchers in this area, Friedrich and Michalak (1983) have systematically considered many of the various forces that may be involved, although the actual data they collected were somewhat limited for this purpose. The logic of the following analysis is modeled somewhat after theirs, and the goal of examining the various sorts of influences that conceivably come into play is similar; however, the procedure here is to piece together relevant data from the many studies already available rather than to gather new data. The attempt is to establish which factors are most likely to mediate the relationship between research productivity and teaching effectiveness when one is found, and which, if any, "explain" it away by being common causes. A concomitant question is whether any evidence exists to show that some of these factors would produce an even larger positive (zero-order) association between research productivity and teaching effectiveness were it not for suppressing or counterbalancing forces. Added to these concerns is a consideration of how varying contexts or conditions affect the relationship between research productivity and teaching effectiveness.

Possible Mediating Variables

Pedagogical Practices and Dispositions

An important class of potential mediating factors between research productivity and overall teaching effectiveness comprises the teacher's actual pedagogical practices and classroom-related dispositions. There is no lack of conjecture about exactly what these elements are and how they come into play. Analysts, either generalizing from their knowledge of pertinent research literature or merely voicing their own views, have suggested a number of instructional dispositions and activities that might be influenced by research productivity, and which in turn presumably affect overall instructional effectiveness. No studies were found that simultaneously related research productivity to faculty members' pedagogical dispositions or practices (as measured either by teachers' own reports or by ratings made by trained observers in the classroom) *and* to overall student-perceived teaching effectiveness. However, there is indirect evidence available if the specific ratings of teachers made by students, discussed earlier, are taken as rough indicators of *actual* classroom-related dispositions, activities, and practices of these teachers that are causally prior to students' assessment of a teacher's overall effectiveness.[13]

Those posited to mediate a positive relationship. It is known that these specific evaluations generally are related positively to overall evaluations (see Feldman, 1976b), and Table 2 of the present analysis has already shown that, across studies, research productivity is in fact related positively to *some* of them. Moreover, the classroom-related dispositions and activities of teachers that have been posited as mediating a positive relationship between research productivity and overall teacher evaluations are easily classifiable into the instructional dimensions already used in Table 2. Thus, it is possible to return to the results presented in that table to see whether research productivity is in fact positively related to the sorts of teacher dispositions and classroom practices that have been assumed (or claimed) to produce a positive association between research productivity and overall instructional effectiveness.

Below are listed seven propositions, extracted from the scholarly literature, about the classroom-related dispositions and activities of teachers affected positively by research productivity. For each of them, citations to the relevant literature are given along with the particular dimension(s) of Table 2 presumed to be the most relevant and the average correlation between research productivity and the dimension(s) (repeated from Table 2).

1. Being productive in research helps teachers not only to keep abreast in their field and gain understanding of the subject matter they teach (cf. Abelson, 1967; Friedrich and Michalak, 1983; Gaff and Wilson, 1971; Linsky and Straus, 1975; Marsh, 1984; McCullagh and Roy, 1975; and McDaniel and Feldhusen, 1970) but also to increase their intellectual vitality and involvement (cf. Abelson, 1967; Friedrich and Michalak, 1983; Harry and Goldner, 1972; Linsky and Straus, 1975; Marsh, 1984; and Stark, 1976): For Dimension No. 3 (Knowledge of Subject), average $r = +.21$ ($p < .001$); and for Dimension No. 4 (Intellectual Expansiveness), average $r = +.15$ ($p = .020$).

2. Research productivity fosters a teacher's own intellectual self-discipline, which may manifest itself in better organization of the course and of classroom lectures (cf. Michalak and Friedrich, 1981; Friedrich and Michalak, 1983) as well as in clearer explanations of course material (cf. Gervetz in Glasman and Killait, 1974, p. 56; Michalak and Friedrich, 1981; and Friedrich and Michalak, 1983) all of which gives students a clearer understanding of where the course is headed and what is expected of them (Friedrich and Michalak, 1983): For Dimension No. 5 (Preparation and Organization), average $r = +.19$ ($p < .001$); for Dimension No. 6 (Clarity and Understandableness), average $r = +.11$ ($p < .001$); for Dimension No. 9 (Clarity of Course Objectives and Requirements), average $r = +.18$ ($p = .002$).

3. Teachers who are productive in research, by thus challenging themselves, in turn expect more of students and challenge them intellectually (cf. Friedrich and Michalak, 1983; also see Michalak and Friedrich, 1981): For Dimension No. 17 (Encouragement of Independent Thought; Intellectual Challenge), average $r = +.09$ ($p = .003$).

4. Teachers who are themselves productive scholars and researchers are more likely to cultivate certain outcomes in students—namely, an ability to reason critically and independently and an enthusiasm for scholarship, systematic inquiry, and research (cf. Abelson, 1967; Bresler, 1968; Friedrich and Michalak, 1983; and McCullagh and Roy, 1975): For Dimension No. 12 (Perceived Outcome and Impact), average $r = +.10$ ($p = .011$).

5. Research productivity, by increasing the teacher's own learning, in-

volvement, and sense of excitement, helps make the teacher more stimulating and interesting to students (cf. Centra, 1983; Friedrich and Michalak, 1983; Harry and Goldner, 1972; Jencks and Riesman, 1968, chap. 12; Linsky and Straus, 1975; and Schmitt, 1965): For Dimension No. 1 (Stimulation of Interest), average $r = +.08$ ($p = .045$).

6. Being productive in research leads to teachers' introducing more relevant material into the classroom (cf. McCullagh and Roy, 1975): For Dimension No. 10 (Value of Course Material), average $r = +.06$ ($p = .039$); and for Dimension No. 11 (Usefulness of Supplementary Material), average $r = +.08$ ($p = .016$).

7. Being productive in research stimulates the teacher's own interest and involvement in the subject matter of the course as well as his or her enthusiasm for teaching (cf. Centra, 1983; Friedrich and Michalak, 1983; Gaff and Wilson, 1971; Harry and Goldner, 1972; Linsky and Straus, 1975; McCullagh and Roy, 1975; and Schmitt, 1965): For Dimension No. 2 (Enthusiasm), average $r = +.09$ ($p = .129$).

With one exception, then, the pertinent data from Table 2 are *consistent with*—and, in this sense, *supportive of*—the various propositions about how research productivity affects pedagogical dispositions and classroom practices of teachers, which themselves are known to be related positively to overall instructional effectiveness. The exception is for the instructional dimension of enthusiasm (as associated with the last proposition), where the relationship is positive but is not statistically significant. Note, however, that the correlations for those instructional dimensions contained in Propositions 3, 4, 5, and 6, though statistically significant, are so small as to have little practical significance. The correlations for the five instructional dimensions that are part of the first two propositions are larger, and thus presumably more likely to be of greater importance as mediating factors than the others. But even the sizes of the correlations for these five (and certainly for the others) imply that the association between productivity and overall teaching effectiveness would be positive but very small (were they the only factors involved in bringing about the association).

Furthermore, the correlations that have been found are consistent with but do not *prove* causation. For example, the largest correlation found in Table 2 is between the faculty member's research productivity and knowledge of the subject matter of the course displayed in the classroom, but this does not automatically mean that the former caused the latter. It may be, say, that the brighter and more intelligent faculty members tend to be both more productive in research and to be more knowledgeable about the subject matter of their courses. Thus the two would be correlated without research productivity having caused the knowledgeability of subject matter (or vice versa).

Those posited to mediate an inverse relationship. The argument has been made by some that research productivity negatively affects certain classroom practices and dispositions of teachers—including some of those just considered as well as certain others listed in Table 2—which dispositions and practices are positively related to overall effectiveness of teachers. If so, these instructional aspects would mediate an *inverse* relationship between research productivity and teaching effectiveness, thereby (in one degree of another) acting as forces that reduce the positive relationship between research productivity and teaching effectiveness.

As a case in point, it might be argued that the more the teacher is involved in specific research projects and committed to being productive in this research, the more likely that he or she is to be narrow in intellectual concerns and to focus on course material that is overly specialized or overly sophisticated for undergraduates (cf. Faia, 1976; Friedrich and Michalak, 1983; McCullagh and Roy, 1975; Michalak and Friedrich, 1971; and Sample, 1972). Yet, empirically, rather than research productivity being inversely correlated with intellectual breadth of the teacher, as just seen, extant research finds research productivity to be *positively* associated with the teacher's intellectual expansiveness. Also as just noted, research productivity is *not* inversely associated with students' ratings of the nature and value, usefulness and relevance of the course material; indeed the correlation is positive though very small. Likewise, the association between research productivity and the closely related dimension of the nature and usefulness of the supplementary material and teaching aids is also positive (though again small).

Michalak and Friedrich (1983) suggest the further possibility that "research with the demands it makes on self-discipline, isolation, and concentration, may do little to enhance the interpersonal skills that seem so important to good teaching—and may even detract from them" (p. 580). In terms of the instructional practices and dispositions as intervening factors, this possibility of detracting is not evident. Although no one of the instructional dimensions listed in Table 2 measures "interpersonal skills" in the classroom in an exact and delimited way, at the very least the following classroom activities and classroom-related dispositions would be involved: Dimensions No. 18 (Respect for Students; Friendliness), No. 16 (Encouragement of Discussion; Openness), and (possibly) No. 8 (Sensitivity to Class Level and Progress). However, as Table 2 shows, research productivity is essentially unrelated to each of these dimensions (average $r = +.05$, $p = .744$; average $r = -.0005$, $p = .537$; and average $r = +.07$, $p = +.244$; respectively).

Other Factors

Other factors besides the teacher's instructional activities and dispositions might mediate the relationship between research productivity and teaching effectiveness, although they have received little attention. Ratz (1975) raises the question of whether research productivity influences teaching assignment, which may, in turn, contribute to differences in teaching effectiveness. For example, if the more highly productive researchers or scholars are given smaller classes to teach or classes that are elective for students—two factors known sometimes to be related positively with teacher and course evaluations (see Feldman, 1978, 1984)—the productivity of these faculty members and their instructional evaluations would tend to be positive as well. However, in a study by Stumpf, Freedman, and Aguanno (1979), research productivity was not significantly correlated with class size or with proportion of required courses taught by the researcher, and controlling for either of these variables did not reduce the association between research productivity and student evaluations. No other studies were found in which such factors were empirically considered when comparing research productivity with teacher effectiveness,

Possible Common Causes

Thus far, the assumption has been that research productivity is causally related to teaching effectiveness, as possibly mediated by certain instructional and other factors (see Figure 1b). Consideration now shifts to whether a positive relationship that is found between research productivity and teacher effectiveness can be explained by these two variables having one or more causes in common (see Figure 1c). If so, the causal interpretation that research productivity affects teaching effectiveness would be a "spurious" interpretation. Moreover, if any common cause(s) related positively to one of the two variables in question and inversely to the other, an inverse relationship between the two would be produced, which, to one degree or another, would counterbalance or suppress any positive relationship between them.

Academic Rank and Age of the Teacher

One "extraneous" variable that might create a so-called spurious relationship between research productivity and teaching effectiveness is the academic rank of the teacher (see Michalak and Friedrich, 1981). If faculty members of higher academic rank are better at both research and teaching, then differences in rank might make research and teaching appear to be causally related (positively) when they really are not. Indeed, it is known from past research that, although rank is not inevitably significantly related to perceived teaching effectiveness, when it is so related, the association is generally positive though weak (see Feldman, 1983). Moreover, at least from those studies with relevant information, at hand for the present review, academic rank is also related positively to research productivity. Across relevant studies, the average correlation between academic rank and *current* research productivity is + .22 (combined $Z = +5.390$; $p < .001$).[14] The average correlation between academic rank and *total career* productivity is + .37 (combined $Z = +12.161$; $p < .001$).[15] It is hardly unexpected that academic rank is correlated positively with total career productivity of the teacher; what may be surprising is that the association is not much higher than that with current productivity.

Academic rank thus qualifies as a possible common-cause variable, such that differences in it might make research productivity and teaching effectiveness appear to be causally related when they really are not. Available evidence suggests that this is not so. Several studies have found that when a statistically significant relationship between research productivity and teaching effectiveness initially exists, it does *not* disappear when academic rank (either alone or along with other variables) is controlled (see Bresler, 1968; Centra, 1983; Grant, 1971; McDaniel and Feldhusen, 1970; Michalak and Friedrich, 1981; Stallings and Singhal, 1970; Stumpf et al., 1979; and Wood and DeLorme, 1976); in Aleamoni and Yimer (1973), Bausell and Magoon (1972), and Voecks (1962), research productivity was not significantly related to teaching effectiveness before or after academic rank was controlled. If a relationship is found, however, it is possible that it diminishes in size under controls rather than disappears. If so, the relationship would be partly spurious. Whether the relationship in fact diminishes is not clear, one way or the other, from existing studies. There is also the possibility of interaction effects of academic rank with research productivity on teaching effectiveness, which will be discussed later in this paper.

Another conceivable influence on both research productivity and teaching effectiveness is the age of the faculty member. From prior research it is known that, in general, the teacher's age is either not related to teaching effectiveness (as assessed by students) or is related inversely to it (for a review of these studies, see Feldman, 1983).[16] Several other studies have found that age, likewise, is inversely associated with research or scholarly productivity (for reviews of these studies, see Blackburn and Lawrence, 1986, and Fox, 1983). However, Blackburn and Lawrence (1986) emphasize that "a larger number of independent studies in diverse settings and across a number of disciplines find a curvilinear relationship between scholarly performance (publications) and age" (p. 275). In the set of studies being reviewed here, only two compared age of faculty members with their research productivity (neither of which was cited in Blackburn and Lawrence, 1986, or Fox, 1983): One found a positive relationship between age and total career productivity (Rossman, 1976), whereas the other found that age was unrelated to current productivity (Clark, 1973).

From these studies taken as a whole, then, it can be seen that age is not inevitably correlated inversely with either research productivity or teaching effectiveness. Yet, given the results of these studies, age might well be inversely associated with both in *some* cases. If so, age could create a "spurious" relationship between the two. Unfortunately, when and where this may have occurred or might be expected to occur—assuming it has or will occur at all—cannot be said, since no studies were found that actually controlled for teacher's age when comparing research productivity with teaching effectiveness.

Individual Attributes and Personality Characteristics of the Teacher

Another set of variables that has been suggested as influencing both research productivity and teaching effectiveness of faculty (and thus creating a spurious relationship between the two) comprises certain of the faculty member's personal attributes or personality characteristics. Ideally such personality attributes would be controlled when relating research productivity to teaching effectiveness, although this has seldom been done in existing studies. Nevertheless, a search among this research can still be made for personality characteristics that are associated with research productivity and also with teaching effectiveness, for these characteristics *potentially* would be common causes of the two variables producing a spurious relationship between them.

One problem here is the lack of research that is *precisely* relevant to the question of whether college teachers of varying personality characteristics also differ in research or scholarly productivity. The only fully relevant research found was that reported by Rushton et al. (1983). In the first of two studies reported by these researchers, 52 psychology professors at the University of Western Ontario were evaluated on 29 personality trait dimensions. A composite criterion of research productivity (called by the researchers "research creativity") was generated from publication and citation counts; and an indicator of teaching effectiveness was created from five years of archival data based on formal student evaluations. In a replication study, a self-report survey form (used to measure both productivity and teaching effectiveness) was sent to 400 faculty members at psychology departments in nine Canadian universities. The following discussion relies heavily on the results of these two companion studies, even though more studies (including research not restricted to any one academic field) will obviously be needed for anything like definitive conclusions.

It should also be pointed out that Fox (1983) cites and reviews several studies on how "productivity levels among scientists" relate to their personality characteristics and attendant attributes. Unfortunately, for present purposes, these studies are not based on samples of college teachers whose research productivity has been measured and then related to their personality traits. Rather, a variety of other methods, samples, and comparison groups have been used, none of which is precisely appropriate for the purposes at hand. This variety includes the construction of composite pictures of the personalities and biographical backgrounds of especially eminent scientists (with no comparison groups composed of less eminent scientists), the study of personality characteristics and productivity of scientists located in different research laboratories and other non-university settings, the comparison of preeminent researchers and scholars with preeminent teachers and administrators, the comparison of scientists with nonscientists, and the like. Thus, the results of these studies, as summarized by Fox (1983), will be considered here merely as suggestive of what more directly relevant studies might show.

Unlike research productivity of teachers as it relates to their personality characteristics, there have been a number of relevant studies comparing student-evaluated teaching effectiveness with these characteristics, and these findings have been systematically inventoried (Feldman, 1986). One difficulty that emerges here is that few personality characteristics of teachers are associated with teaching evaluations when these characteristics are measured by teachers' responses to self-report personality inventories or their own self-descriptions on questionnaires, whereas many personality characteristics have been found to be associated with teaching evaluations when these characteristics are measured by students' or colleagues' perceptions of the teachers. These latter associations are hard to interpret, however, for they may contain certain artifactual or confounding elements (see Feldman, 1986, for a fuller discussion). For purposes of the present analysis, then, greatest confidence will be placed on those associations that are found between teachers' personality characteristics and teaching effectiveness for *both* ways of determining personality characteristics.

General ability. The first attribute to be considered is general ability—the suggestion being that some part of the positive relationship between research productivity and teaching effectiveness is due to the fact that faculty members of superior ability are more productive scholars or researchers and at the same time are more effective teachers (Black, 1972; Centra, 1983; Faia, 1976; Friedrich and Michalak, 1983; Harry and Goldner, 1972; Linsky and Straus, 1975; Michalak and Friedrich, 1981; Stark, 1976; and Stumpf et al., 1979). Although the exact meaning of general ability is unspecified in these sources, reference seems to be to some general level of intelligence as measured, say, by one or another established, standardized test of adult intelligence. Assuming this to be correct, no study was found in which intelligence, so measured, was controlled when comparing research productivity of faculty members with their teaching effectiveness. Considering research productivity alone, Fox (1983) reports in her review that "among scientists with doctoral degrees, measured intelligence (IQ) is, in fact, very high. But, within this already select group, IQ correlates very weakly with productivity and achievement in science" (p. 288). Whether or not faculty members' scores on IQ-type tests also vary with students' evaluations of their teaching is not currently known.

One study—by Wood and DeLorme (1976), of 69 faculty members at the College of Business Administration at the University of Georgia (1971–1973)—did find that an association between research productivity of the teachers and their teaching evaluations held when controlling on what the authors called a proxy variable for faculty ability. This proxy variable was "developed from a quantified evaluation of each professor by his department chairman" (p. 78). Since the association remained, faculty ability (at least as measured) could not account for the positive relationship. However, it may partially explain the relationship that was found in this particular study, for, in a regression analysis, the beta coefficient for research productivity was much smaller when the ability measure was in the regression than when it was not.

One methodological problem in this study is that the measure of general ability was confounded with the other two variables, since the department chairmen rated ability of faculty members' knowing something about their research productivity and their success as teachers. A measure of general ability that has been established independently of these other two would obviously be preferred. Of course, a measure of a more specific ability pertinent to research performance (cf. Cole and Cole, 1973, pp. 248–249) *and* to instructional effectiveness may be needed rather than a measure of some general ability. Or it may turn out that research ability must be distinguished from teaching ability, which would require a whole different model of causal influences (see, for example, Marsh, 1984, and Note 19 of the present analysis).

Intelligence and intellectual curiosity. Closely connected to measures of general ability, presumably, are items or scales—either on personality inventories or in self-report questionnaires—that measure "intelligence" (referring to a person's brightness, quickness, and cleverness), and "intellectual curiosity" (referring to a person's reflectiveness, intellectuality, and cultural and aesthetic sensitivity). Although these sorts of personality characteristics were found by Rushton et al. (1983) to be positively associated with research productivity in one of their two studies of psychology teachers, they were not found to be so in the second. Moreover, across other relevant studies, variables such as these have not generally been found to be related to student-perceived teaching effectiveness when they were measured by faculty members' own responses to personality inventories and self-description questionnaires (although they have been associated with teaching effectiveness when these traits were measured by students' and colleagues' perceptions) (see Feldman, 1986). Given this pattern of results, "intelligence" (as brightness) and intellectual curiosity seem not to be strong candidates as common causes of research productivity and teaching effectiveness capable of producing a spurious relationship between them, although clearly more research is needed (including research that actually controls for these variables when relating research productivity to teaching effectiveness) before they can be definitely eliminated.

Responsibleness, persistence, and orderliness. Michalak and Friedrich (1981) have suggested that being organized and self-disciplined might be personality traits that create positive relationships between research productivity and teaching effectiveness by positively influencing each of them. In Feldman (1986), the relevant cluster of personality characteristics was labeled "responsible, conscientious, persistent and orderly." It was found in that review that traits such as these, when measured by teacher's self-descriptions and responses to personality inventories, were *not* related to teacher evaluations (although they were when measured by students' and colleagues' perceptions of them). In the first of their two studies, Rushton et al. (1983) also found that such traits related positively to research productivity when they were measured by students' and colleagues' perceptions but not when measured by teachers' own responses to personality inventories and self-descriptions. Likewise, with some exceptions, these sorts of characteristics tended not to be related to research productivity in their second study (using self-descriptions only). By contrast, in her review of research on the personality characteristics of scientists—research somewhat tangential to the present analysis it will be remembered—Fox (1983) found some evidence that highly productive scientists tended to be highly persistent, to prefer precision and exactness, and to show strong control of impulses. No studies were located for the present analysis in which these particular traits were actually controlled when comparing research productivity of teachers with their teaching effectiveness.

Ascendancy, forcefulness, and leadership. At this point, it is worth looking more closely at the overall results across the two studies reported by Rushton and his associates (1983), for an additional personality factor that might account in part for a positive relationship between research productivity and teaching effectiveness is suggested. These investigators summarize the relevant portion of their findings as follows:

> we may characterize the creative [i.e., productive] researcher as ambitious, enduring, seeking definiteness, dominant, showing leadership, aggressive, independent, not meek, and non-supportive . . . [and] the effective teacher as liberal, sociable, showing leadership, extraverted, low in anxiety, objective, supporting, non-authoritarian, not defensive, intelligent, and aesthetically sensitive. . . . In addition to the replicated traits [across the two studies], one or [the] other of the studies found the researcher also to be low in sociability, intelligent, curious, compulsive, orderly, not seeking of help, not fun loving, authoritarian, defensive, and non-neurotic. Similarly, other traits found for the effective teacher included fun loving, changeable, low in harm avoidance, low in neuroticism, intellectually curious, enduring, orderly, attention seeking, ambitious, non-impulsive, and approval seeking. It is interesting to note that the constellations of traits defining the creative [productive] researcher and the effective teacher are approximately orthogonal. While one cluster suggests independence, achievement orientation, dominance, and striving to create cognitive order, the other denotes an easier-going, intelligent liberality. The only trait that effective researchers and teachers shared in common was leadership. The one on which researchers were opposite was supportingness, with researchers being low and teachers high. (Rushton et al., 1983, pp. 110–111)

Note that a personality factor these researchers clearly found to be a possible common cause of research productivity and teaching effectiveness was leadership. However, across relevant studies, Feldman (1986) did not find that the cluster of personality traits in which leadership was placed in his study ("ascendancy, forcefulness, conspicuousness and leadership") was related to teacher evaluations when these traits were measured by personality inventories of teachers or their own self-descriptions (although they were when measured by colleagues' and students' perceptions of the teacher's personality). Also, no other studies were found that compared the personality variable of leadership or related traits of college teachers with their research productivity,[17] so it is not known whether the positive association found by Rushton et al. (1983) is a general one.

Supportiveness, tolerance, and warmth. The passage excerpted from Rushton et al. (1983) also raises the important possibility that certain personality variables to some degree "suppress" an otherwise positive relationship between the research productivity of faculty members and their perceived teaching effectiveness. In this particular instance, the personality characteristic suggested was "supportingness"—defined by the authors as a personality disposition to give sympathy and comfort, to be helpful, to be indulgent—which was inversely associated with research productivity but positively associated with teaching effectiveness.

No other studies were located in which college teachers' trait of supportingness was compared with their scholarly productivity, although the Fox (1983) review of tangential studies did point to evidence that more highly productive scientists tend to be "persons who are not overly concerned with other persons' lives" and to show "a preoccupation with ideas and things rather than people." As for the positive association between supportingness and teacher evaluations found in Rushton et al. (1983), there is clear and consistent corroboration in other studies. Thus in the review by Feldman (1986), the relevant trait-cluster, called "positive view of others: sympathetic, tolerant, supportive, and warm," was positively related to teaching effectiveness when it was measured by use of personality inventories and teachers' self-descriptions (across 10 studies, average $r = +.15$, combined $Z = +3.901$, $p < .002$) as well as when it was measured by using the perceptions of students and colleagues (across 7 studies, average $r = +.53$, combined $Z = +8.986$, $p < .001$).

Given the set of findings by Rushton et al. (1983) and Feldman (1986), together with the review by Fox (1983), the possibility remains that the trait of supportingness does indeed "suppress" the degree to which research productivity is positively associated with teaching effectiveness. Whether this possibility can be claimed as an actuality can only be determined by future studies that systematically focus on this trait and related characteristics and control for them when relating research productivity to teaching effectiveness.

Sociableness and extraversion. The particular cluster of traits involving supportingness has similarities to, but should not be confused with, such traits as sociability (vs. unsociability) and extraversion (vs. introversion), which also have been speculated to be related inversely to research productivity but positively to teaching effectiveness (cf. Friedrich and Michalak, 1983; Linsky and Straus, 1975; and Michalak and Friedrich, 1981).[18] Evidence that this is so is much less clear and consistent than in the case of supportingness. In the research of Rushton et al. (1983), there is a little evidence that sociability was inversely associated with research productivity, but only in the first of their two studies. (The Fox review, 1983, does point to a bit of evidence that productive scientists tend to be "isolated from social relations.") As for teaching effectiveness, in the review by Feldman (1983) the trait cluster of "sociable, gregarious, friendly and agreeable" was found not to be related to teacher evaluations when these traits were measured by personality inventories and self-descriptions (although, once again, when measured by colleagues' and students' perceptions, these traits of the teachers were positively associated with teaching effectiveness).

Time and Effort

Another factor thought to affect both teaching and research is the actual effort of time spent on one relative to the other (cf. Blackburn and Clark, 1975; Braunstein and Benston, 1973; Friedrich and Michalak, 1983; Gaff and Wilson, 1971; Harry and Goldner, 1972; Hoyt, 1974; Linsky and Straus, 1975; Michalak and Friedrich, 1981; Sample, 1972; and Stark, 1976). By itself, this factor would produce an inverse relationship between research productivity and teaching effectiveness, thus reducing any positive relationship produced by other forces. As Michalak and Friedrich (1981) have put it: "The more time and effort a faculty member devotes to research—presumably increasing research productivity—the less time he or she has to devote to teaching—presumably detracting from teaching effectiveness . . ." (p. 149).[19] This contention might seem obviously true and without need of defense. However, the surprise here is that extant research is supportive for only certain parts of the proposition, as will be elaborated in the following discussion of pertinent research.

Time and effort spent in research and teaching as correlates of research productivity. Some evidence exists for thinking that research productivity is in fact positively influenced by the time and effort spent in research (and, conversely, negatively affected by the time and effort spent in teaching activities), although this evidence is correlational in nature, and somewhat mixed besides. Hayes (1971) reports that time or effort spent in research (estimated by the faculty members' department head) is positively and significantly related to number of publications, but an exact correlation is not given. Hoyt and Spangler (1976) report a correlation of +.84 (for the social-behavioral sciences) and +.76 (for natural-mathematic sciences) between the faculty members' time commitment and research accomplishments at the University of Kansas. It must be noted, however, that in this study the department heads both appraised the significance of the faculty members' research efforts and estimated how much time was devoted to these efforts, so the correlations are most probably artificially inflated.

Indeed, a faculty member's own estimate of the number of hours per week spent in research activities, while still positively correlated with research productivity is less strongly and less consistently so. Thus McCullagh and Roy (1975), studying faculty at Appalachian State University, report a correlation of +.46 ($p < .001$) between number of hours per week faculty said they spent in preparation of academic materials intended for publication and number of academic articles published, but a statistically insignificant correlation of +.12 for number academic books published. Harry and Goldner (1972) found a positive but statistically insignificant correlation ($r = +.19$) between number of hours per week faculty "at a large public urban midwestern university" said they spent in research and their number of publications.

One other piece of research is of particular interest here. Studying 86 professors in 23 departments in natural, mathematical, medical, and biological sciences at the University of Missouri–Columbia, Jauch (1976) found that the individuals who spent more time in research tended to be higher on most of the 11 research performance measures in his study (e.g., rho between time spent in research and publication = +.404, $p < .001$; for citations, rho = +.206, $p < .05$; for a grant efficiency index, rho = +.311, $p < .05$; and for a total performance index across the 11 measures, rho = +.463, $p < .001$). It was also the case, conversely, that those who spent more time in teaching-related activities—classroom teaching, class preparation, grading, individualized student help, etc.—performed less well on these same measures of research productivity (e.g., for publication, rho = −.378, $p < .001$; citations, rho = −.067, not statistically significant; grant efficiency index, rho = −.217, $p < .05$; total performance index, rho = −.265, $p < .01$). Jauch notes that "as to why time is allocated as it is, a faculty member may either get more satisfaction from either of these activities (and thus devote more time to it), or he may feel that more time must be spent in order to receive favorable rewards (salaries and promotions)" (p. 8). From his research, he found that both of these were operating, but that satisfaction seemed more important than potential reward. The individual who spent more time in research got more feelings of accomplishment from research than from teaching activities, just as those spending more time in teaching got more feelings of accomplishment from teaching than from research activities.

Time and effort spent in research and teaching as correlates of teaching effectiveness. As just seen, time and effort spent in research activities often correlate with research productivity (and with teaching effectiveness) in ways expected. It is the reverse side of the coin that produces the surprises. Following from the results of the research just reviewed, especially that of Jauch (1976), it would be expected that amount of time and energy spent in research activities would be inversely related to teaching effectiveness, just as the time and effort spent in teaching-related activities would be expected to be positively associated with teaching effectiveness. But little of the available evidence supports either of these expectations. Table 3 offers brief summaries of studies relating time spent in research activities to overall effectiveness of teachers as perceived by students. Note that Grant (1971) did find an inverse association of perceived teacher effectiveness with percentage of time allocated to nonsponsored (nonfunded) research and writing, but not with percentage of time allocated to sponsored (funded) research. In each of the three other studies given in Table 3, although the correlation between the variables under consideration was inverse, it was not statistically significant. Moreover, across these studies, the average correlation is only −.07; and this correlation, too, is not statistically significant (combined $Z = -1.046$; $p = .296$).

TABLE 3. Summary of Results of Studies Relating Time or Effort Spent in Research Activities to Overall Effectiveness of Teachers as Perceived by Students

(*) *Bausell and Magoon (1972):* Percentage of time spent in research × overall instructor evaluation: $r = -.11^*$; $Z = -1.120^*$.

Grant (1971): As the percentage of time allocated by faculty members to nonsponsored (nonfunded) research and writing increased, the courses they taught received lower mean scale scores on the "recommended" item of an evaluation form completed by students. For sponsored (funded) research, no statistically significant relationship was found. In both cases, r or its equivalent cannot be determined from information given.

(*) *Harry and Goldner (1972):* Weekly hours spent in scholarly (research activities × percentage of students in the class giving the instructor an "A" (overall evaluation of his or her performance): $r = -.04^*$; $Z = -0.347^*$.

(*) *McCullagh and Roy (1975):* Average hours per week spent in preparation of academic material intended for publication × general rating of classroom teaching effectiveness: $r = -.05^*$; $Z = -0.354^*$.

Note. The general note for Table 1 also applies to this table.

Other research results are also relevant here, if not quite as directly so. If time and effort devoted to research is inversely related to overall teaching effectiveness, then certain of the more specific instructional activities previously discussed would be expected to come into play, in certain ways, as mediating variables (see Figure 1d). It might be expected, for example, that the more time and effort that is spent in research, the less likely instructors are to be organized and carefully prepared for class (Instructional Dimension No. 5 in Table 2), the less likely they are to give timely feedback of high quality to students (Instructional Dimension No. 15), and the less likely they are to be available to students and helpful to them in terms of advising and consultation (Instructional Dimension No. 19) (cf. Blackburn and Clark, 1975; Gaff and Wilson, 1971; McCullagh and Roy, 1975; and Friedrich and Michalak, 1983). In short, time spent in research would be expected to be inversely associated with these variables, each of which is known to be positively associated with overall teaching effectiveness.

Unfortunately, only one study (Grant, 1971) could be found in which time spent in research activities was compared with students' ratings of teachers on specific instructional dimensions. In this research, the percentage of time spent in nonfunded research and writing indeed was inversely related to student ratings of the "preparation" and "responsiveness" items on an evaluation form, but percentage of time spent in funded research was unrelated to both evaluation items. The "preparation" item clearly represents Instructional Dimension No. 5. It is not clear, however, whether the "responsiveness" item represents Instructional Dimension No. 19 (Availability and Helpfulness), No. 16 (Encouragement of Discussion; Openness), or No. 18 (Respect for Students; Friendliness), for information about the items (other than their names) is not given in the report. Although there are data on other specific evaluations in this particular study, none of them represents Instructional Dimension No. 5 (Feedback to Students).

There is another way of determining whether the amount of time or effort devoted to research is inversely related to teaching effectiveness through the mediation of one or more of these three instructional dimensions. If time or effort spent in research is a common cause affecting research productivity positively and each of the three instructional dimensions under consideration negatively, other things equal, this would lead to a (spurious) inverse correlation between research productivity and each of these three instructional dimensions. But as Table 2 shows, and as previously discussed in a different context, on average (across relevant studies) students' ratings of the frequency and quality of feedback from the teacher (Instructional Dimension No. 15) and of the availability and helpfulness of the teacher (Instructional Dimension No. 19)[20] are unrelated to research productivity rather than being inversely related; and students' evaluation of teachers on preparation and organization (Instructional Dimension No. 5) is positively related to research productivity.

There is almost no support, then, for the proposition that time or effort devoted to research is inversely related to teaching effectiveness either in some direct way or indirectly through its negative effects on teacher's preparation and organization, quality of the teacher's feedback to students, or the teacher's helpfulness and availability. Surprisingly, even the amount of time or effort devoted to *teaching* and closely related activities does not seem much related to teaching effectiveness, at least in studies located for the present review, as seen in Table 4. Across the 8 studies with correlations (of the 12 studies summarized in Table 4), the average correlation between time spent by teachers in teaching activities and students' overall evaluations of these teachers is barely positive (average $r = +.0013$), although statistically significant (combined $Z = +1.993$; $p = +.046$).

This somewhat anomalous situation of statistical significance comes about because two of the five studies with positive correlations have very large samples of faculty members but very small correlations (in Delaney, 1976, $r = +.08$, $N = 3{,}558$; and in Pezzullo, Long, and Ageloff, 1976, $r = +.05$, $N = 1{,}930$), whereas the three with inverse correlations have small sample sizes (see Clark, 1973; Harry and Goldner, 1972; and Wood, 1978). An average r does not take into account variation in the size N of the sample on which the r is based, whereas the combined Z is *partially* dependent on the N of each study (because these individual N's in part are reflected in the

TABLE 4. Summary of Results of Studies Relating Time or Effort Spent in Teaching or Teaching-Related Activities to Overall Effectiveness of Teachers as Perceived by Students

---

(*) *Bausell and Magoon (1972):* Percentage of time spent in preparation for teaching × overall instructor evaluation: $r < .10$ counted as $r = .00$, $Z = 0.000$; and for percentage of time spent counseling students, $r = +.10$, $Z = +1.017$; average $r = +.05$*, $Z = +0.508$*, $N = 105$.

Centra and Creech (1976): Teachers with teaching loads of 13 or more semester hours were rated higher than any of the other four groups of teachers (those with teaching loads of 3 or fewer hours, 4 to 6 hours, 7 to 9 hours, and 10 to 12 hours. Those with 10–12 hours received the second highest ratings, although this was not appreciably different from the ratings of those with smaller teaching loads. A 1 × 5 analysis of variance was statistically significant, but results cannot be converted to an r. When teaching assistants were excluded from the analysis, and only two levels of teaching load were considered (less than 10 semesters and 10 or more semester hours), the results failed to reach statistical significance, suggesting little difference in ratings of the more permanent staff regardless of teaching load. Again, results cannot be converted to an r.

(*) *Clark (1973):* Credit-hour teaching load × overall rating of instructor's teaching effectiveness: $r = -.25$*; $Z = -1.665$*; $N = 45$.

(*) *Delaney (1976):* Teaching load × "Student-Teacher Relationship" Factor Scale of the *Student Instructional Report:* $r = +.11$, $Z = +6.579$; and for "Course Objectives and Organization" Factor Scale, $r = +.05$, $Z = +2.983$; average $r = +.08$*, $Z = +4.778$*, $N = 3{,}558$.

Goldsmid, Gruber, and Wilson (1977): Winner and runner-ups for superior teaching award taught more courses than did a control group of faculty (see p. 438). Also, course load was related to how highly ranked the nominee was (see Table 3), although direction of results is not given. In neither case can r or its equivalent be determined from the information given.

Grant (1971): The percentage of an instructor's total time spent in classroom instruction (including course preparation and advising students), the time allocated to specific courses, and the amount of time spent in activities supporting instruction (such as program supervision, counseling students, and committee work) were all unrelated to the "recommended" item of an evaluation form completed by students; r or its equivalent cannot be determined from information given.

(*) *Harry and Goldner (1972):* Weekly hours devoted to teaching (class time, lectures, preparation, grading, seeing students, etc.) × percentage of students in the class giving the instructor an "A" (overall evaluation of his or her performance); $r = -.18$*; $Z = -1.592$*; $N = 79$.

(*) *Hoffman (1984b):* Ancillary teaching activities (such as program development,

TABLE 4. ***(Continued)***

participation in field experiences, field projects, off-campus instruction, supervision of independent study students and interns) × student evaluations of teachers: $r = -.12^*$; $Z = -0.956^*$; $N = 65$.

Lasher and Vogt (1974): Instructional work load with respect to amount of course preparations per quarter was related to the letter grade students assigned the teacher (overall evaluation), but instructional load with respect to number of sections taught per quarter was not; in neither case was the direction of the association given, nor can r or its equivalent be determined from the information given.

(*) McDaniel and Feldhusen (1970): Number of lecture hours × composite instructor evaluation: $r = .04$, $Z = +0.344$; for number of laboratory hours, $r = +.18$, $Z = +1.561$; and for number of counseling (student advising) hours, $r = +.23$, $Z = +2.005$; average $r = +.15^*$, $Z = +2.197^*$, $N = 76$.

(*) Pezzullo, Long, and Ageloff (1976): Teaching load × effectiveness of instructor compared to other instructors student has had: $r = +.05^*$; $Z = +2.197^*$; $N = 1{,}930$.

(*) Wood (1978): Number of student credit hours for regular classes over the prior three rating years × three-year average general student evaluation: $r = +.04$, $Z = +0.183$; for typical number of undergraduate advisees, $r = +.39$, $Z = +1.854$; for typical number of graduate advisees, $r = +.25$, $Z = +1.163$; for total number of theses committees, $r = +.26$, $Z = +1.212$; and for student credit hours gained by teaching nonassigned courses (over the prior four years), $r = +.20$, $Z = +0.926$; average $r = +.23^*$, $Z = +1.068^*$, $N = 23$.

Note. The general note for Table 1 also applies to this table.

size of the individual Z's). Weighting both the individual r's and Z's by N (specifically, *df*, or degree of freedom, which is $N-2$), and using the method of adding weighted Z's to calculate combined Z (see Rosenthal, 1984, No. 5 on pp. 94–95), produces a somewhat larger positive average correlation across the studies ($r = +.06$) and a much larger combined Z ($+5.232$, $p < .001$). It could be argued that using this weighting procedure in a sense gives too much importance to (that is, overweights) the correlation of $+.08$ in the Delaney (1976) study (with its Z of $+4.778$). Even so, the average association is still extremely small in size and of little practical significance.[21]

Time and effort: Additional findings and overview. If time and effort devoted to research were strong determinants of both research productivity (positive) and teaching effectivenc⁻ (negative), ones that outweighed other forces, an *inverse* relationship between research productivity and teaching effectiveness would be expected. However, as established at the outset of this paper, almost none of the zero-order associations between research productivity and teaching effectiveness in various studies have been inverse. An assumption, often made only implicitly, in the prediction of an inverse relationship between these two variables is that time or effort spent in research activities is negatively related to time or effort spent in teaching activities. In a study of faculty members of the Colleges of Liberal Arts and Education at a large, public urban midwestern university, Harry and Goldner (1972) did find an inverse relationship between time spent in each set of activities, but it was not all that strong ($r = -.31$). As these investigators made clear, increments of time spent on research and scholarly activities were associated with only small decrements on time or effort devoted to teaching. They further presented indirect evidence that increments of time spent by faculty in research or scholarly activity came less from decrements in time in other work-related activities (whether time spent in teaching, committee work, consulting, private practice, or the like) and more from decrements in time spent in leisure activities and with family or friends.

If these findings from this particular study hold more generally, the lack of a strong inverse relationship between time spent in research and in teaching might account (in part) for why the observed association between research productivity and teaching effectiveness is not negative. Of course, it would still be possible for time or effort spent in research to affect research productivity positively and teaching effectiveness negatively—if only moderately or even weakly so—which to some extent would reduce or suppress the otherwise more strongly positive relationship between research productivity and teaching effectiveness by other forces. If so, this would help account for the weak positive correlations that have been observed in most of the studies. But even this possibility seems not to be the actual case. From evidence in several studies (as previously reviewed), time or effort spent in research, while often related positively to research productivity, appears to be unrelated to teaching effectiveness (rather than inversely related). Further, although the more time spent in teaching activities the more likely is research productivity to be adversely affected (at least from the little bit of relevant evidence that could be found), teaching effectiveness again does not appear to be particularly affected (rather than positively affected) by effort devoted to teaching.[22]

Possible Contexts and Conditions

It is possible that the relationship between research productivity and teaching effectiveness is contingent upon certain circumstances. That is, there may be conditions or situations under which the relationship is larger or smaller, or even reversed in direction. If so, although the general association between research productivity and teaching effectiveness may be weak across various conditions or contexts, it may be much stronger in some of them. The search, then, is for factors that "specify" the original relationship (see Cole, 1976; and Hyman, 1955) or, alternatively put, for factors that have statistically significant "interaction effects" with research productivity on teaching effectiveness.

Career Stage of Teacher

One such factor is the stage of the faculty member's career, as indicated by his or her academic rank. Although academic rank was not found in the present analysis to be a common-cause variable creating a spurious association between research productivity and teaching effectiveness, it is still possible that the strength (and even the direction) of the association between the two *varies* by the academic rank of the teacher. Thus, it has been suggested either that the positive relationship between research productivity and teaching effectiveness is weaker or that the relationship between the two is negative for teachers in the earliest stages of their career (assistant professors and/or nontenured faculty), whereas the positive relationship is not only positive but progressively stronger for teachers at more advanced stages of their career (higher-ranked, tenured faculty) (cf. Centra, 1983; Harry and Goldner, 1972; and Michalak and Friedrich, 1981).

Underlying this proposition is the thought that faculty who are obliged structurally to engage in research and scholarly activity may concentrate on research at the expense of teaching in order to improve their chances of winning tenure (or being promoted). These professionally young teachers may find that their research and scholarly projects detract from their teaching, or at least go less hand-in-hand with it, whereas professionally older or more established faculty members through the passage of time may have become more adept at doing the kinds of research and scholarly activities that enhance their teaching. Or, if the proposition holds true, a different explanation may account for the results: Those teachers who are less adept at juggling their research and teaching may leave (or be dropped) from the faculty, whereas those who have learned to reduce any incompatibilities between the two (and even enhance any complementarities) may be disproportionately retained (cf. Goldsmid, Gruber, and Wilson, 1977).

Despite the reasonableness of both of these rationales, no consistent support across studies was found for the proposition. In the first of two studies reported by Centra (1983), the expectation that research productivity and teaching effectiveness would be more strongly correlated for teachers in the middle or later years of their teaching career than in the early years was supported for social science teachers, but not for teachers in the natural sciences or humanities. In the second study, a larger positive relationship in the later years was evident neither for social sciences nor for natural sciences and humanities, although it was for the category of "professional areas" (engineering, business, education, and health professions) that was added to this study. From tables in Bresler (1968), it can be seen that the positive association between research support and teacher evaluation was higher for the senior faculty (compared to junior faculty) for the sciences and engineering, and humanities, faculty; however, just the opposite was true for the social science faculty.

In comparing nominees for a teaching award with a control group of faculty, Goldsmid et al. (1977) found that for full professors, some 86% of the nominees had as good or better a publication record than their counterparts in the control group, whereas this percentage dropped to 54% for associate professors and 57% for assistant professors. This result does offer some support for the proposition under consideration, although greater support would have shown the percentage of 57% for associate professors to be closer to the 86% for full professors. Neither Harry and Goldner (1972) nor Stallings and Singhal (1970) found a statistically significant interaction effect between tenure-nontenured status of the teacher and research productivity on student-perceived teaching effectiveness. Voecks (1962) found no association between research productivity and perceived teaching effectiveness for either full professors or associate and assistant professors (grouped together). Finally, in direct contrast to the proposed pattern of results, Friedrich and Michalak (1981) found that the positive relationship between research productivity and teaching effectiveness actually decreased as rank of the teacher increased. Given the wide variation in results across these several studies, it is no surprise that when meta-analytic procedures are applied to them, the results do not support the proposition that research productivity is stronger for individuals who are at more advanced stages of their teaching career.[23]

Academic Discipline

Research productivity might be associated with teaching effectiveness in some disciplines and not others. Several reasons for this have been suggested, including the following: the possibly differential ease with which teachers in different disciplines can adopt the content of their research (with its particular level of abstraction and difficulty) to the undergraduate classroom, general differences among disciplines as to faculty members' collaboration with students in doing research; differences among teachers in different disciplines with respect to availability of extramural support and funding for research and the consequent impact on teaching; and possible differences in the "sets" that students bring to various disciplines, which affect the likelihood that teachers will fulfill students' expectations (see especially Hoyt and Spangler, 1976; and Michalak and Friedrich, 1981).

Another important reason for comparing research productivity with teacher effectiveness within academic disciplines is the possibility that an indicator of research productivity (whether the number of publications, citation counts, or some other measure) may not, in fact, provide a uniform index of research accomplishment because of differences in the meaning of research and in style of publication for different disciplines. To the extent that this is a problem, analyzing the relevant association between research productivity and teaching effectiveness *within* disciplines—even if only within broad disciplinary areas—is a step toward a solution.

Table 5 summarizes the results of the few studies that have compared the associations between research productivity and teaching effectiveness for the three broad academic divisions of humanities, social sciences, and natural sciences. Two of the studies added a category of "professional areas," and results for them have also been included.[24] Using the data in this table, the average correlation for humanities is +.22 (combined $Z = +4.540$; $p < .001$), for social sciences, +.20 (combined $Z = +4.851$; $p < .001$); for natural sciences, +.05 (combined $Z = -0.218$; $p = .827$); and for professional areas (with data for only two studies) $r = +.06$ ($Z = +1.973$; $p = .048$). Considering only these results, it would seem that research productivity and teaching effectiveness for humanities and social sciences are more strongly related (positively) than they are in professional areas (although the correlation for the latter is based on only two studies), and that the two are unrelated for natural science faculty. Assuming this indeed to be the case, one possible explanation is suggested by Michalak and Friedrich (1981), who hypothesize that "research in the natural sciences, in contrast to research in the social sciences and humanities, may be at a level of abstraction and complexity that renders it of little utility in the classroom" (p. 593).

Yet, conclusions in this area should not be drawn too quickly. Results in Hoyt and Spangler (1976), not given in Table 5, found a positive relationship between research involvement and students' perceived progress on "professional skills, viewpoints," "discipline's methods," "thinking, problem solving," and "personal responsibility" for teachers of natural sciences, whereas the relationship was inverse for teachers in the social science. It is true that these results are for what is considered here to be a specific instructional dimension (Perceived Outcome or Impact of Instruction, Dimension No. 12) rather than an overall evaluation of instruction (which is why the study was excluded from Table 5). Yet this particular dimension is one of the two or three dimensions that past research has shown to be the most highly related

TABLE 5. Summary of Results of Six Studies Comparing the Associations between Research Productivity and Student-Perceived Teaching Effectiveness in Different (Broadly Defined) Academic Areas

Humanities

(*) Braunstein and Benston (1973): For Humanities faculty (departments not given): r (rho) = +.24*; Z (for N estimated at 347/5, or 69.4) = +1.999*.

(*) Bresler (1968)[a]: For Arts and Humanities faculty (classics, drama and speech, English, fine arts, German, music, philosophy, religion, and the Romance languages): r = +.21*; Z (based on t of 0.85) = +.0845*.

(*) Centra (1983, Study 1): For Humanities faculty (departments not given): average (weighted) r (across 3 categories of years of teaching experience) = +.13*; Z = +4.623*.

(*) Centra (1983, Study 2): For Humanities faculty (departments not given): average (weighted) r (across 3 categories of years of teaching experience) = +.02*; Z = +0.311*.

(*) Friedrich and Michalak (1981): For Humanities faculty (departments not given): r (for research merit rating) = +.48*; Z = +2.374*.

Voecks (1962): For Letters faculty (drama, English, languages, philosophy, and speech), chi-square indicates no statistically significant relationship. Data provided are insufficient for calculating an r or its equivalent, and the direction of the association cannot be determined.

Social Sciences

(*) Braunstein and Benston (1973): For Social Science faculty (departments not given): r (rho) = −.05*; Z (for N estimated at 347/5, or 69.5) = −0.411*.

(*) Bresler (1968)[a]: For Social Science faculty (child study, economics, education, government, history, and sociology): r = +.11*; Z (based on t of 0.44) = +0.439*.

(*) Centra (1983, Study 1): For Social Science faculty (departments not given): average (weighted) r (across 5 categories of years of teaching experience) = +.13*; Z = +4.524*.

(*) Centra (1983, Study 2): For Social Science faculty (departments not given): average (weighted) r (across 3 categories of years of teaching experience) = +.23*; Z = +3.625*.

(*) Friedrich and Michalak (1981): For Social Science faculty (departments not given): r (for research merit rating) = +.57, Z = +2.862; r (for number of citations) = +.54, Z = +2.442. Average (weighted) r = +.56*; average (weighted) Z = +2.671*.

TABLE 5. *(Continued)*

Voecks (1962): For Social Science faculty (anthropology, business administration, the Far East, history, political science, psychology, and sociology), chi-square indicates no statistically significant relationship. Data provided are insufficient for calculating an *r* or its equivalent, and the direction of the association cannot be determined.

Natural Sciences

(*) *Braunstein and Benston (1973):* For Natural Science faculty (departments not given): *r* (rho) = −.04*; *Z* (for *N* estimated at 347/5, or 69.4) = −0.329*.

(*) *Bresler (1968)*[a]*:* For Science and Engineering faculty (biology, chemical engineering, chemistry, civil engineering, geology, mathematics, mechanical engineering, physics, and psychology): *r* = +.28*; *Z* (based on *t* of +1.34) = +1.340*.

(*) *Centra (1983, Study 1):* For Natural Science faculty (departments not given): average (weighted) *r* (across 3 categories of years of teaching experience) = −.04*; *Z* = −0.901*.

(*) *Centra (1983, Study 2):* For Natural Science faculty (departments not given): average (weighted) *r* (across 3 categories of years of teaching experience) = −.06*; *Z* = −1.076*.

(*) *Friedrich and Michalak (1981):* For Natural Science faculty (departments not given): *r* (for research merit rating) = +.07, *Z* = +0.421; *r* (for number of citations) = +.20; *Z* = +1.125. Average (weighted) *r* = +.13*; average (weighted) *Z* = +0.784*.

Voecks (1962): For Natural Science faculty (botany, chemistry, engineering, geology, mathematics, meteorology, pharmacology and medicine, physics, and zoology), chi-square indicates no statistical significant relationship. Data provided are insufficient for calculating an *r* or its equivalent, and the direction of the association cannot be determined.

"Professional Areas"

(*) *Braunstein and Benston (1973):* For Engineering faculty, *r* (rho) = +.36, *Z* (for *N* estimated at 347/5, or 69.4) = +3.050; for Management faculty, *r* (rho) = −.31, *Z* (for *N* estimated at 347/5, or 69.4) = −2.606. Average *r* = +.03*; average *Z* = +0.222*.

Bresler (1968): No data given. (Data for Engineering faculty cannot be extracted from data for Natural Science and Engineering faculty).

Centra (1973, Study 1): No data given for this category.

(*) *Centra (1973, Study 2):* For faculty in engineering, business, education and health professions, average (weighted) *r* (across 3 categories of years of teaching experience) = +.09*; *Z* = +2.568*.

TABLE 5. ***(Continued)***

Friedrich and Michalak (1981): No data given for this category.

Voecks (1962): No data given for this category.

Note. The general note for Table 1 also applies to this table.
[a]The note *within* the entry for Bresler (1968) in Table 1 also applies to the entries for Bresler (1968) in this table.

to overall evaluation (see Feldman, 1976b). So it is probable that if Hoyt and Spangler (1976) had used an overall evaluation item (or items), the results would be quite similar to what they did find (i.e., a positive association for natural sciences and an inverse association for social sciences).

This would mean that if the results from their study were averaged in with those in Table 5, as a proxy for the results had an overall evaluation of the teacher been measured, the average positive association for natural sciences (across studies) would increase in size (possibly becoming statistically significant), while the average positive association for social sciences (across studies) would decrease in size. If so, humanities would have the highest (average) positive correlation, followed by the social sciences and the natural sciences, with some possibility that the associations for the latter two would end up not far apart in size. Unfortunately, the exact correlations from the Hoyt and Spangler (1976) study cannot be averaged in – even using perceived progress as a substitute for an overall evaluation item – because the correlations that are needed are not given in the study, and the information that *is* given in the study (including the relevant figure on p. 119) is not sufficient for obtaining or computing a correlation.

A further complexity in incorporating the results from the Hoyt and Spangler (1976) study into the meta-analysis is that they are not altogether comparable to those in the studies summarized in Table 5. In Hoyt and Spangler, research involvement was measured by combining time spent in research (research commitment) and research productivity (accomplishment). In the present analysis, the two have been separated, with the latter considered as the dependent variable of interest and the former a "common-cause" variable. The degree to which combining them may have produced results different from those if only research productivity (accomplishment) had been used is unknown. Moreover, unlike the other studies in Table 5, the results in the Hoyt and Spangler study are adjusted for students' initial motivations in taking each course. Still, for all these incompatibilities, the results in this study by Hoyt and Spangler create just enough "doubt" to make conclusions based on Table 5 extremely tentative. Information from another study or two added to the table might well change the overall results and conclusions drawn.

Type of College or University

There may well be other contexts that specify the relationship between research productivity and instructional effectiveness, but research is very sparse. One possibility is the type of college or university. Linsky and Straus (1975) wonder if the role of researcher and that of teacher are "related differently in high prestige and in other schools" (p. 101) or whether the relationship between research productivity and teaching effectiveness varies "between universities with graduate schools and small undergraduate college" (p. 101), but they present no data to answer these questions. Likewise, Kulik and Kulik (1974) speculate that the nature and degree of the relationship "probably is different at different schools" (p. 54). Michalak and Friedrich (1981), having found a positive correlation between research productivity and teaching effectiveness at a small liberal arts college, conjecture that "it might in fact be the case that involvement in research at small liberal arts colleges . . . influences teaching more than it does at large universities" (p. 593). These investigators realize, of course, that the data they present cannot be taken as evidence one way or another, for they have not done comparative research across different types of educational institutions.

Faia (1976) has done the necessary comparative research. He found that the positive association between a faculty member's having published and receiving an award for outstanding teaching was stronger for "institutions weak on research emphasis" (comprehensive universities and colleges not offering the doctorate, liberal arts colleges, and two-year colleges and institutes) than it was for "institutions strong on research emphasis" (research universities and doctoral-granting universities). Further research along these lines seems well warranted. As part of replicating and extending research in this area, it would be particularly useful to break down institutions of higher education into more than two categories as well as to have different indicators of productivity and teaching effectiveness than those used by Faia.[25] It should be remembered that even if additional research confirms that positive correlations between research productivity and teaching effectiveness are substantially higher at certain colleges and universities, this does not necessarily show that research productivity at these institutions is causally beneficial to teaching at these schools. It may merely mean, for example, that common causes are more likely to create spurious correlations between research productivity and teaching effectiveness at these institutions. Appropriate analyses in future research could determine whether or not this is so.

SUMMARY AND CONCLUSIONS

Strong opinions about the connection between research (or scholarship) and teaching, based on informal observation alone, are not hard to find. Writing in *School and Society* in 1958, George B. Cutten was quite certain from his experiences in academe that "the more research a professor has done, the more books and articles he has written, the better teacher he is supposed to be. But the opposite is more likely to be the case" (p. 372). In direct response to this contention, Lewis Leary (1959), with equal certainty, proclaimed: "The popular image of the scholar as pedant immersed in library or laboratory has about the same validity as the popular image of Mr. Chips, Miss Dove, or Mark Hopkins and his log. . . . The fact is that our best teachers are almost without exception our best scholars. . . . Scholarship is not at a different pole from teaching" (p. 362). Extant research supports neither position. The present review found that, on the whole, scholarly accomplishment or research productivity of college and university faculty members is only slightly associated with teaching proficiency.

Nearly without exception, the many studies located for review have not found (for their total samples) statistically significant *inverse* associations between research productivity or scholarly accomplishment of faculty, as measured in a variety of ways, and students' assessments of these teachers' overall instructional effectiveness. Rather, these relationships almost always have been in the positive direction, although, more often than not, they have been statistically insignificant. Excluding studies where results were not reported in the form of product-moment correlations or whose results could not be converted to such correlations, the average correlation for the remaining 29 studies was calculated to be +.12. This correlation is statistically significant when individual probability levels from the separate studies are combined. This small positive association holds when research productivity is measured by publication counts, by indicators of research support, and by ratings from others (such as department chairpersons), but not when measured by citation counts. This last indicator of research productivity comes the closest of all the indicators to measuring the actual *quality* of the productivity and accomplishments, which thus appears to be unrelated to teaching effectiveness.

An obvious interpretation of these results is either that, in general, the likelihood that research productivity actually benefits teaching is extremely small or that the two, for all practical purposes, are essentially unrelated. In either case, an important conclusion would be that productivity in research and scholarship does not seem to detract from being an effective teacher. Before making any interpretations or drawing any conclusions, however, it is important to know whether underlying the small positive correlations (statistically significant or otherwise) are larger positive and negative forces that are more or less counterbalancing one another. For example, perhaps research productivity does have fairly strong negative effects on teaching effectiveness but they are somewhat more than offset by factors creating a positive relationship between the two. To learn more about what might be behind the zero-order associations, existing research was analyzed in terms of factors that potentially mediate the relationship between research productivity and teaching effectiveness and those that potentially are common causes of both.

Considering first the factors that might mediate positive relationships between research productivity and teaching effectiveness, a commonly suggested causal sequence is that research productivity positively affects certain classroom practices and pedagogical dispositions of teachers, which, in turn, positively influence their overall instructional effectiveness. Using meta-analytical procedures, positive associations between research productivity and many of these practices and dispositions (themselves known to relate positively to teaching effectiveness) were indeed found. But most of them were very small. The largest of the correlations (a little over or under +.20) were between research productivity and knowledge of the subject, intellectual expansiveness, preparation and organization, and clarity of course objectives and requirements. Much smaller average (positive) correlations were found between research productivity and clarity and understandableness, perceived impact of the course, encouragement of independent thought and intellectual challenge, stimulation of interest, and the value, relevance, and usefulness of the course material selected.

The characteristics of classes or courses that faculty members of varying productivity in research and scholarship are assigned to teach constitute another sort of potential mediating factor. For example, if the more highly productive researchers or scholars are given smaller classes to teach or courses that are elective for students—both of which have sometimes been found to be related positively with student evaluations of teachers and courses—the productivity of these faculty members and their overall instructional evaluations would also be positive. Little can be said beyond noting this possibility, however. Only one study was found that had relevant data, and in it neither class size nor electivity of courses turned out to be a mediating factor.

As for common causes that might create an observed positive relationship between research productivity and teaching effectiveness, even if the two are not causally related, academic rank of the faculty member was considered first. Meta-analysis of relevant data, from studies collected for the present review, showed that the academic rank of the faculty member was positively related to both "current" and total-career productivity. Because some (not all) studies have also found academic rank to be positively related to teaching effectiveness, it is possible in certain instances that the positive correlation observed between research productivity and teaching effectiveness is due mainly to their both being dependent in some way on the academic rank of the faculty member. This appears *not* to be the case, however; for the statistically significant positive association found between research productivity and teaching effectiveness in several studies remained when academic rank was controlled. (Whether the original associations became smaller cannot be told from these studies.) Also unlikely as a common cause is the age of the faculty member, for age has been found to be associated (inversely) with research productivity and with student evaluations in some studies but not others. No studies were located in which the faculty member's age was controlled when comparing research productivity with teaching effectiveness, so it is unknown to what degree and in what circumstances it might account for positive associations that are found between them.

A number of analysts have suggested that differences in general ability among faculty members may explain the positive relationship under consideration. That is, faculty members who are generally superior in ability tend to excel in both research and teaching, thus explaining the positive association between the two. Although this argument seems plausible, neither the meaning of ability in this regard (apart from general intelligence) nor how it is to be measured has been well specified in the literature. In the one research study that included a measure of the general ability of faculty members, research productivity and teaching effectiveness were still related positively after it was controlled, although not as strongly. However, the measure of ability (chairpersons' ratings) was not an especially satisfactory or convincing one.

Finally, certain aspects of the faculty member's personality conceivably might affect positively both research productivity and instructional effectiveness. Put into clusters of personality traits, they are (1) intelligence (in the sense of brightness, quickness, and cleverness) and intellectual curiosity (reflectiveness, intellectuality, and cultural and aesthetic sensitivity); (2) responsibleness, persistence, and orderliness; and (3) ascendancy, forcefulness, and leadership. In relevant research that has been done using these variables, there is no *consistent* evidence across studies of positive associations of any of them with either research productivity or student-rated teaching effectiveness, although scattered positive associations can be found.

In all, none of the factors discussed are clear-cut cases of common causes of research productivity and teaching effectiveness. Even if some were, they, along with the set of mediating factors that were found, are clearly weak in strength. By themselves, they would produce only a small positive correlation between research productivity and teaching effectiveness. Because only a small average (positive) correlation was in fact found between research productivity and teaching effectiveness in the present analysis, little "room" is left for the operation of counterbalancing negative forces that would suppress what would otherwise be larger positive relationships between research productivity and teaching effectiveness.

The implication of this logic is sustained by available evidence. Thus, the classroom practices and pedagogical dispositions of teachers that are sometimes said to be negatively affected by research productivity – thereby mediating an inverse relationship between this productivity and teaching effectiveness – in fact are *not* related to research productivity. High producers of research, compared with low producers, are no less likely (nor any more likely) to be friendly in class, to show concern for students, to encourage discussions, to be open to others' opinions, or to be sensitive to class level and progress. Nor are they more likely – indeed, they may be a little less likely – to be intellectually narrow and to assign course material that is either overly specialized or overly sophisticated for students.

Many people, both within and outside of academia, are convinced that the time or effort spent by faculty members in research, though obviously beneficial to their research productivity and scholarly accomplishments, detracts from their instructional effectiveness. In effect, this reasoning considers the amount of time or effort spent in research as a common cause that creates an inverse relationship between research productivity and teaching effectiveness by being related in opposite directions with each. Existing evidence does not support this contention, however. Although the data do support the proposition that the more time spent in research the greater the likelihood of high research productivity (and one study showed the obverse, that time spent in teaching was inversely related to research productivity), time or effort spent in research is *not* negatively related to student-perceived teaching effectiveness as thought. (For that matter, neither was time and effort faculty members devoted to teaching and teaching-related activities found to be particularly related to students' evaluations of them.)

One cluster of personality traits was found to relate in opposite ways to research productivity and teaching effectiveness. With some consistency across studies, supportiveness, tolerance, and warmth (not to be confused with sociableness and extroversion) were associated inversely with research productivity but positively with teaching effectiveness. Whether, and how, these associations actually affect the relationship between research productivity and teaching effectiveness is unknown, however, for no study was located in which one or another of these traits was controlled when comparing research productivity with teaching effectiveness.

It should be clear at this point that there is an advantage of having investigated which factors are or are not related to research productivity and teaching effectiveness (and what happens when these factors are controlled). Either a lack of correlation or a small positive correlation between the two variables now means more than it otherwise would. Confidence is increased in the working assumption that, when research productivity and teaching effectiveness are found to be unrelated, they are essentially independent of each other (rather than that relatively strong forces that would otherwise create a large positive relationship, say, are being cancelled out by forces of the same strength that by themselves would produce a large negative relationship). Similarly, if a weak positive relationship is found, the immediate assumption would not be that strong positive forces are outweighing negative forces to some slight degree, but that only weak forces are in operation. As discussed, increases in research productivity may slightly raise the likelihood that certain pedagogical practices beneficial to effective teaching occur, although, because of insufficient data, the possibility cannot yet be ruled out that the small positive relationship merely represents the fact that faculty of superior ability tend to be good at both research and teaching. If there are any counterbalancing negative factors, they too are weak. It is not clear what these factors are, although the set of personality traits of supportiveness, tolerance, and warmth is a possibility because these traits are associated in opposite ways with research productivity and teaching effectiveness.

The existence and strength of the relationship between research productivity and teaching effectiveness no doubt vary by circumstances or condition. Indeed, there may be discoverable contexts in which positive associations can be expected routinely to occur, or to be larger rather than smaller. Based on what little evidence now exists, career stage of the faculty member is *not* one of these conditions, at least not in any consistent way across studies. There is some evidence, however, that positive associations between research productivity and instructional effectiveness are more likely to occur, and to be larger, within the humanities and the social sciences than in the natural sciences, although this conclusion is extremely tentative. Also, one study found somewhat larger positive correlations at colleges weaker in research emphasis. It is conceivable that two, three, or even more contexts or conditions may combine to produce higher-level interaction effects (cf. Centra, 1983). Perhaps for certain departments, academic divisions or disciplines at certain schools (and even then perhaps only for faculty at certain career stages), research productivity may actually affect teaching effectiveness negatively, just as there may be certain other schools and disciplines within them that particularly promote much larger positive relationships between the two than are generally found. What these specific conditions or contexts are, or whether they even exist in reliable and specifiable ways, awaits future research.

Despite the fact that the present review[26] is based on a rather comprehensive collection of previous studies that have been analyzed intensively, uncertainties in drawing conclusions have obviously not been altogether eliminated. There are other ways of determining research productivity, academic accomplishment, and overall instructional effectiveness than the ones used in the studies reviewed here, and perhaps research using them would produce different results and lead to different conclusions. Some of the studies have methodological weaknesses, which makes their findings problematic. Moreover, there are significant gaps in the research literature; on more than one occasion, important information was simply not available.

The present review is not without its problems either. In some cases, meta-analytic procedures were used with a smaller number of studies than is ideally desirable. In more than one instance, indirect and somewhat unsatisfactory indicators had to be relied upon in the absence of better ones. As only one example, in analyzing specific classroom practices of faculty members (as possible mediating factors between research productivity and overall teaching effectiveness), it would have been preferable to have the judgments of trained observers in the classroom as measures of these practices rather than (or, at least, in addition to) students' perceptions of them. Furthermore, the present analysis often had to draw together disparate and scattered sets of data in an overly piecemeal fashion, indicating, incidentally, that the area of research under review quite definitely needs more multivariate analyses, with causal modeling, of complete sets of relevant variables (see Stumpf et al., 1979, and Friedrich and Michalak, 1983, for beginnings of such analyses and models). Still, for all the difficulties and uncertainties, it has been possible to discern some relatively coherent patterns of findings in existing research and to draw some tentative conclusions from them that can help guide future research.

NOTES

1. Not included in this table is a piece by Frumkin and Howell (1954), in which one "effective" teacher is compared with one "ineffective" teacher. More than the small sample size led to the exclusion of this study. Although the researchers selected each of the two teachers from a student-nominated pool of teachers, they do not report the criteria for selection. Unknown, therefore, is the extent to which the differences that are found between the research productivity of the two teachers is due merely to an arbitrary selection by the investigators rather than the sample of two representing some larger population of effective and ineffective teachers. Also excluded from Table 1 are the following studies, which sometimes are claimed to have data relevant to the topic at hand, but whose indicators do not really measure both research productivity and student-assessed teaching effectiveness (as conceived here and by the investigators whose research is included in the table): Braxton (1983); Maslow and Zimmerman (1956); McGrath (1962); and Woodburne (1952). A brief review of most of these excluded studies can be found in Linsky and Straus (1975); also see Faia (1976).
2. When possible, Z's were calculated for an individual association by using the t of F values

given in the study and a formula supplied by Rosenthal (1984, p. 107, no. 5.15) for estimating Z knowing t (or F, which can be converted to t). If more than one association pertained to a study, the individual Z's were averaged. Where neither t's nor F's were given in the study, t's were estimated by using the summary r's (either a single correlation or an average correlation) and formula 2.3 in Rosenthal (1984, p. 107), from which Z's could then be computed.

3. Various methods exist for combining the probability levels associated with separate results in order to generate an overall probability relating to the existence of the relationship. Rosenthal (1978, 1984) discusses several such methods of combining independent probabilities to get an overall estimate of the probability that the separate p levels would have been obtained were the null hypothesis of no relationship true in each of the cases. Adding up the separate Z's and dividing the resultant sum by the square root of the number of studies perhaps is the simplest and most routinely applicable of the methods, and is the one used here.
4. It might be thought that the combined Z is significant primarily because of the very large component Z ($+25.408$) contributed by the Faia (1976) study, this component Z owing its size in part to the very large number of faculty studied ($N = 53{,}034$). However, even if one considers this component Z to be zero rather than $+25.408$, the combined Z is still $+7.588$ ($p < .001$).
5. Not surprisingly, various indicators of research productivity have been found to intercorrelate positively with one another (among the studies reviewed in the present analysis, see Bresler, 1968; Dent and Lewis, 1976; Plant and Sawrey, 1970; and Rushton, Murray, and Paunonen, 1983; for some other studies, see Cole and Cole, 1973; Meltzer, 1956; and Schrader, 1978). The present analysis does not explore the advantages of using one of these procedures of determining productivity over the others. Discussion of the merits and demerits of one or another indicator of research productivity and scholarly accomplishment can be found in Blackburn (1974), Blackburn and Lawrence (1986), Clark (1957), Cole and Cole (1973), Crane (1965), Dent and Lewis (1976), Garfield (1979), Harmon (1963), Myers (1970), Price (1963), Rushton, Murray, and Paunonen (1983), Smith and Fiedler (1971), and Wilson (1964).
6. In Table 1, see Aleamoni and Yimer (1973), Centra (1973, Study 1), Centra (1973, Study 2), Clark (1973), Faia (1976), Freedman, Stumpf, and Aguanno (1979), McCullagh and Roy (1975), Rushton, Murray, and Paunonen (1983), Wood and DeLorme (1976), and Wood (1978).
7. In Table 1, see Bausell and Magoon (1972), Dent and Lewis (1976), Harry and Goldner (1972), Hicks (1974), Linsky and Straus (1975), Marquardt, McGann, and Jakubaukus (1975), McDaniel and Feldhusen (1970), Siegfried and White (1973), Stallings and Singhal (1970, Study 1), Stallings and Singhal (1970, Study 2), and Stavridis (1972).
8. Four other studies in Table 1 (McDaniel and Feldhusen, 1970; Plant and Sawrey, 1970; Ratz, 1975; and Teague, 1981) contain information about the relationship between research support and teacher evaluation, but they could not be included in the meta-analysis because of insufficient data. In general, these four do *not* find a statistically significant relationship between research support and teacher evaluations. This does not mean, however, that the statistically significant average correlation of $+.17$ that was found for the two studies would be reduced if correlations could be gotten from these other four studies and averaged in. The exact change in average correlation (as well as the combined probability level) would depend on the size of each correlation in combination with the size of the sample. (For example, there might be relatively large correlations based on small samples that would not be statistically significant; these would not reduce the average correlation found, and the combined probability level might well be statistically significant.)

9. One study (Marsh and Overall, 1979) used the faculty member's self-rating on a scale of "scholarly productivity," and this too produced a positive correlation with teacher evaluations of $+.14$, which just misses statistical significance at the .05 level ($Z = +1.893$; $p = .058$). In Hoffman (1984b), where a mixture of research productivity and professional associations was used (number of publications, grants received, attendance at and participation in meetings, awards, prizes, and special recognition by the professional community), the relevant correlation was negative ($r = -.25$; $Z = -2.015$; $p = .044$). The studies, when broken down by type of indicator of research productivity, add to more than 29 because some of them used more than one type of indicator (whose results were combined in the present analysis when calculating the overall average association across the studies).
10. These findings are not as discrepent with Frey (1978) as it might seem, for Frey fails to note that the inverse correlation of $-.23$ he reports between the "rapport" scale and research productivity is not statistically significant for an N of 42, and so the two variables might better have been treated as unrelated. Incidentally, Frey's division of specific evaluation items into "pedagogical skill" and "rapport" scales is roughly analogous to the clustering of the specific instructional dimensions of the present analysis into the two large clusters that Feldman (1983, 1984) found useful. The first cluster (Dimension No.'s 1–12) involve the teacher's task of presenting material in his or her role of *actor* or *communicator,* as Widlak, McDaniel, and Feldhusen (1973) have put it, while the second cluster (Dimension No.'s 13 and 15–19) have to do with the teacher's task of facilitating students' involvement in the class and in learning, as part of the teacher's role of *interactor* or *reciprocator* (using once again the names suggested by Widlak and his associates, 1973). Averaging across the average correlations found for each dimension (as given in Table 2), weighting by the N for each dimension, produces an average correlation of $+.15$ ($Z = +9.936$; $p < .001$) for the first cluster of dimensions and an inconsequential (though statistically significant) correlation of $+.02$ ($Z = +2.244$; $p = .025$) for the second cluster.
11. Conceivably a positive correlation could be due to the positive effects of teaching effectiveness on research productivity, although this seems to be a less plausible interpretation (cf. Black, 1972; Centra, 1983; Jencks and Riesman, 1968, p. 533; Linsky and Straus, 1975; and Michalak and Friedrich, 1981).
12. In this case, any correlation found between research productivity and the particular set of intervening variables (just as between research productivity and teaching effectiveness) is produced by the common-cause variables.
13. Students' perceptions of teachers' pedagogical dispositions and practices may be inaccurate, of course, and may not even always be causally prior to their overall evaluations of teachers, because of the so-called halo effect of overall evaluation on specific evaluations (see Feldman, 1976b, for a discussion of this latter possibility). Yet there is some evidence that the specific assessments of teachers made by students along various instructional dimensions do correlate positively with the presumably more "objective" and unbiased assessments made by trained observers (see Cranton and Hillgartner, 1981; Halstead, 1972; Love et al., 1977; Love, Sandoval, and Cohen, 1978; Stallings and Spencer, 1967; Touq, 1972; Touq and Feldhusen, 1974; and Tracey and Tollefson, 1979).
14. This result is based on data in four studies, as follows: Aleamoni and Yimer (1973), average $r = +.32$ and $Z = +6.226$ (for number of publications for 1966–1969 unweighted, and weighted in two different ways, $r = +.32$, $+.32$, $+.33$, respectively); Clark (1973), $r = +.22$ and $Z = +1.460$ (for number of publications in the prior five years); Freedman, Stumpf, and Aguanno (1979), $r = +.16$ and $Z = +1.815$ (for number of publications in a three-year period); and Hoffman (1984b), $r = +.16$ and $Z = +1.278$ (for research productivity for the calendar year).
15. This result is based on data in five studies, as follows: Dent and Lewis (1976), average

$r = +.43$ and $Z = +4.215$ (for total number of publications, $r = +.59$, for number of citations by colleagues within the instructor's own discipline, $r = +.41$, and for number of citations by scholars outside the instructor's discipline, $r = +.30$); for Linsky and Straus (1975), average (weighted) $r = +.34$ and average (weighted) $Z = +10.623$ (for number of publications weighted by type of publication over approximately a twenty-year period, $r = +.39$, $Z = +13.190$, $N = 1.065$; and for number of citations, $r = +.24$, $Z = +5.766$, $N = 563$); Rossman (1976), $r = +.61$ and $Z = +7.184$ (for measure of research productivity based on personnel folders, using number and type of publications); Stallings and Singhal (1970, Study 1), $r = +.26$ and $Z = +2.968$ (for total number of publications weighted by type of publication), and Stallings and Singhal (1970, Study 2), $r = +.20$ and $Z = +2.203$ (for total number of publications weighted by type of publication).

16. See Feldman (1983) for a discussion of the apparent paradox between academic rank being positively correlated with perceived teaching effectiveness but age, which itself is positively correlated with academic rank, being inversely correlated with perceived teaching effectiveness.
17. In her review of studies tangential to the present analysis, Fox (1983) did find some evidence that productive scientists show "personal dominance."
18. A distinction that also should be made explicit here is that between (1) sociability, extraversion, and related attributes as a set of general personality traits of the individual that can be measured, say, by scales in personality inventories and (2) the "interpersonal skills" teachers display and use in the classroom as part of their pedagogical practices, which presumably involve such instructional dimensions as teachers' respect for students and friendliness toward them, encouragement of class discussion and openness to students' opinions, and (possibly) sensitivity to class level and progress. These were the instructional practices that were discussed in an earlier section of this paper as possible mediators of an inverse relationship between research productivity and overall teaching effectiveness, and found not to be.
19. Marsh (1984) proposes incorporating this time or effort factor into a more complex model that includes the abilities of the faculty and the reward structure of the school. He posits two different abilities that vary among faculty members, which are presumed to correlate positively with each other. These are: (1) research ability, which is seen to be positively associated with research effectiveness; and (2) teaching ability, which is seen to be positively associated with teaching effectiveness. If abilities were the only aspects to consider, the positive relationship between these two abilities would itself produce a positive relationship between research productivity and teaching effectiveness. However, other factors are involved. Research ability, as mediated by the external and intrinsic rewards to be gained from doing research, affects the amount of time spent on research and, consequently, research effectiveness, just as teaching ability, as mediated by the external and intrinsic rewards to be gained for teaching, affects the amount of time spent in teaching-related activities and consequently, teaching effectiveness. Because time spent on research presumably is inversely correlated with time spent on teaching, the positive relationship between research and teaching effectiveness that would be produced by the positive relationship between research and teaching abilities is weakened. No observed relationship between the two, or, at best, a weak positive relationship, results.
20. Note that Friedrich and Michalak (1983), not having direct indicators of time and effort spent by faculty in research and scholarly activities, use in their research three specific evaluation items as substitute indicators. Two of the items ask about feedback from the teacher, the other about the teacher's availability outside class.
21. It may be noted here that time spent in other professional activities by the faculty member also appears to be unrelated to teaching effectiveness across studies. Thus, based on data in

Clark (1973), McCullagh and Roy (1975), and Wood (1978), involvement in teachers' workshops, conferences, and professional associations (measured by either time spent in these activities or the number of them in which a faculty member participates) has an average correlation of + .05 (combined $Z = +.317$; $p = .751$) with perceived teaching effectiveness. Involvement in consulting activities is not associated with teaching effectiveness either across the three studies with relevant data: average $r = +.08$ and combined $Z = +0.822$ ($p = .411$) (see Bausell and Magoon, 1972; McCullagh and Roy, 1975; and Wood, 1978). Finally, involvement in academic committees and other administrative work also appears to be unrelated to student-perceived teaching effectiveness; across seven studies with pertinent data, average r is + .01 and the combined Z is +0.666 ($p = .505$) (Bausell and Magoon, 1972; Clark, 1973; Marquardt, McGann, and Jackubauskas, 1975; McCullagh and Roy, 1975; McDaniel and Feldhusen, 1970; Siegfried and White, 1973; and Wood, 1978).

22. Focusing only on those studies reviewed here rather than on all research on work in academe and the professional performance and activities of faculty members, it would be of interest to see whether improvements or refinements of measurement would bring any differences in research findings here. In addition to measuring teaching and research effort by absolute number of hours spent in each, it might be useful to develop indices of time spent in one of them *relative* to the other, as well as relative to the total amount of time spent in professional activities (of which teaching and research are only a part) in comparison with leisure, family, personal, and related activities. Moreover, work could be done on measuring commitment to research or to teaching other than by indexing the mere amount of time spent in doing one or the other of them. This is not to say that the studies reviewed for the present analysis have not begun such attempts – for example, some studies measure proportion or percentage of time devoted to research or to teaching rather than absolute number of hours spent in either of them – but more could be done.
23. Excluding Stallings and Singhal (1970), Goldsmid, Gruber, and Wilson (1977), and Voecks (1962), in which r or its equivalent was not given and could not be calculated from the available information, the relevant associations presented in each of the studies just reviewed were divided into those found for (1) assistant professors, or nontenured faculty, or faculty with less than six years of teaching and (2) associate and full professors, or tenured faculty, or faculty with more than six years of teaching. Averaging across the product-moment correlations (or their equivalents) for categories of academic disciplines when necessary (in Bresler, 1968, and Centra, 1983), the average correlation between research productivity and student-perceived teaching effectiveness across studies for the first group of teachers (in the early stages of their career) was compared with the average correlation for the second group of teachers (in the middle and late stages of their career). These correlations turned out to be just about the same ($r = +.15$ vs. $r = +.17$), thus showing in another way the lack of support for the proposition being "tested."
24. Not included in the table are studies that collected data on teachers within one of these disciplines, but where no comparisons across disciplines were made; see Dent and Lewis (1976), whose data are for five fields within the social sciences; Frey (1978), for five fields within the natural sciences; Hoffman (1984b), for education; Marsh and Overall (1979), for the social sciences; Rushton, Murray, and Paunonen (1983), for psychology; Siegfried and White (1973), for economics; and Wood (1978), for education.
25. Linsky and Straus (1975) have written that "if teaching and research are uncorrelated for . . . [a] population of relatively 'high powered' institutions, then it is very unlikely that they would be correlated in other institutions in which research is not a major goal" (p. 95). But, assuming the results in Faia (1976) can be replicated and extended, just the opposite seems to be true.
26. The thoughtful reviews and analyses by Crimmel (1984), Finkelstein (1984, pp. 120–127),

and Webster (1986) were located too late for some of the interesting ideas in them to be included in the present analysis. They are recommended as useful additions to the reviews that have been more fully incorporated into the present paper.

REFERENCES

Abelson, P. H. (1967). Justifying academic research. *Science* 157: 759.

Ahern, J. (1969). Characteristics of teachers who have received an "outstanding teacher" award from New England institutions of higher education in the five year period beginning with the academic year 1963–64. Unpublished doctoral dissertation, University of Massachusetts.

Aiken, L. R., Jr. (1975). Procedures and problems in designing a college course evaluation questionnaire. *College and University* 50(3): 247–253.

Aleamoni, L. M., and Yimer, M. (1973). An investigation of the relationship between colleague rating, student rating, research productivity, and academic rank in rating instructional effectiveness. *Journal of Educational Psychology* 64(3): 274–277.

Aleamoni, L. M., and Yimer, M. (1974). Graduating senior ratings relationship to colleague rating, student rating, research productivity and academic rank in rating instructional effectiveness. Research Report No. 352. Urbana-Champaign, Ill.: Measurement and Research Division, Office of Instructional Resources, University of Illinois.

Bausell, R. B., and Magoon, J. (1972). The validation of student ratings of instruction: An institutional research model. Newark, Del.: College of Education, University of Delaware.

Black, S. (1972). Interactions between teaching and research. *Universities Quarterly* 26(3): 348–352.

Blackburn, R. T. (1974). The meaning of work in academia. In J. I. Doi (ed.), *Assessing Faculty Effort* (New Directions for Institutional Research No. 2), pp. 75–99. San Francisco: Jossey-Bass.

Blackburn, R. T., and Clark, M. J. (1975). An assessment of faculty performance: Some correlates between administrator, colleague, student, and self ratings. *Sociology of Education* 48(2): 242–256.

Blackburn, R. T., and Lawrence, J. H. (1986). Aging and the quality of faculty job performance. *Review of Educational Research* 56(3): 265–290.

Braunstein, D. N., and Benston, G. J. (1973). Student and department chairman views of the performance of university professors. *Journal of Applied Psychology* 58(2): 244–249.

Braxton, J. M. (1983). Teaching as performance of scholarly-based course activities: A perspective on the relationship between teaching and research. *Review of Higher Education* 7(1): 21–33.

Bresler, J. B. (1968). Teaching effectiveness and government awards. *Science* 160: 164–167.

Centra, J. A. (1983). Research productivity and teaching effectiveness. *Research in Higher Education* 18(4): 379–389.

Centra, J. A., and Creech, F. R. (1976). The relationship between student, teacher, and course characteristics and student ratings of teacher effectiveness. PR-76-1. Princeton, N.J.: Educational Testing Service.

Clark, K. E. (1957). *America's Psychologists: A Survey of a Growing Profession.* Washington, D.C.: American Psychological Association.

Clark, M. J. (1973). Organizational stress and professional performance among faculty members at a small college. Unpublished doctoral dissertation, University of Michigan.

Cole, J. R., and Cole, S. (1973). *Social Stratification in Science.* Chicago: University of Chicago Press.

Cole, S. (1976). *The Sociological Method* (2nd ed.). Chicago: Rand McNally.

Cornwell, C. D. (1974). Statistical treatment of data from student teaching evaluation questionnaires. *Journal of Chemical Education* 51(3): 155–160.

Crane, D. (1965). Scientists at major and minor universities: A study in productivity and recognition. *American Sociological Review* 30(5): 699–714.

Cranton, P. A., and Hillgartner, W. (1981). The relationship between student ratings and instructor behavior: Implications for improving teaching. *The Canadian Journal of Higher Education* 11(1): 73–81.

Crimmel, H. H. (1984). The myth of the teacher-scholar. *Liberal Education* 60(3): 183–198.

Cutten, G. B. (1958). The college professor as teacher. *School and Society* 86: 372–375.

Delaney, E. L. (1976). The relationships of student ratings of instruction to student, instructor and course characteristics. Paper read at the annual meeting of the American Educational Research Association. (ERIC Document Reproduction Service No. ED 126 140)

Dent, P. L., and Lewis, D. J. (1976). The relationship between teaching effectiveness and measures of research quality. *Educational Research Quarterly* 1(3): 3–16.

Faia, M. A. (1976). Teaching and research: Rapport or mésalliance. *Research in Higher Education* 4(3): 235–246.

Feldman, K. A. (1976a). Grades and college students' evaluations of their courses and teachers. *Research in Higher Education* 4(1): 69–111.

Feldman, K. A. (1976b). The superior college teacher from the students' view. *Research in Higher Education* 5(3): 243–288.

Feldman, K. A. (1977). Consistency and variability among college students in rating their teachers and courses: A review and analysis. *Research in Higher Education* 6(3): 223–274.

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education* 9(3): 199–242.

Feldman, K. A. (1979). The significance of circumstances for college students' ratings of their teachers and courses. *Research in Higher Education* 10(2): 149–172.

Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education* 18(1): 3–124.

Feldman, K. A. (1984). Class size and college students' evaluations of teachers and courses: A closer look. *Research in Higher Education* 21(1): 45–116.

Feldman, K. A. (1986). The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: A review and synthesis. *Research in Higher Education* 24(2): 139–213.

Finkelstein, M. J. (1984). *The American Academic Profession: A Synthesis of Social Scientific Inquiry Since World War II.* Columbus, Ohio State University.

Fox, M. F. (1983). Publication productivity among scientists: A critical review. *Social Studies of Science* 13(2): 285–305.

Freedman, R. D., Stumpf, S. A., and Aguanno, J. C. (1979). Validity of the Course-Faculty Instrument (CFI): Intrinsic and extrinsic variables. *Educational and Psychological Measurement* 39(1): 153–159.

Frey, P. W. (1978). A two-dimensional analysis of student ratings of instruction. *Research in Higher Education* 9(1); 69–91.

Friedrich, R. J., and Michalak, S. J., Jr. (1983). Why doesn't research improve teaching? Some answers from a small liberal arts college. *Journal of Higher Education* 54(2): 145–163.

Frumkin, R. M., and Howell, W. H. (1954). Effective college teaching: A preliminary causal analysis. *Journal of Educational Research* 47(9): 683–693.

Gaff, J. G., and Wilson, R. C. (1971). The teaching environment. *AAUP Bulletin* 57(4): 475–493.

Gage, N. L. (1978). *The Scientific Basis of the Art of Teaching.* New York: Teachers College Press.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics* 1: 359–375.

Glasman, N. S. and Killait, B. R., eds. (1974). *Second UCSB Conference on Effective Teaching.* Santa Barbara, Calif.: Graduate School of Education and Office of Instructional Development, University of California, Santa Barbara.

Glass, G. V., McGaw, B., and Smith, M. L. (1981). *Meta-Analysis in Social Research.* Beverly Hills, Calif. Sage.

Goldsmid, C. A., Gruber, J. E., and Wilson, E. K. (1977). Perceived Attributes of Superior Teachers (PAST): An inquiry into the giving of teacher awards. *American Educational Research Journal* 14(4): 423–440.

Grant, C. W. (1971). Faculty allocation of effort and student course evaluations. *Journal of Educational Research* 64(9): 405–410.

Halstead, J. S. (1972). Students' ratings of college classroom verbal interaction as related to ratings of instructor teaching effectiveness. Unpublished doctoral dissertation, Purdue University.

Hammond, P. E., Meyer, J. W., and Miller, D. (1969). Teaching versus research: Sources of misperceptions. *Journal of Higher Education* 40(9): 682–690.

Harmon, L. R. (1963). The development of a criterion of scientific competence. In C. W. Taylor and F. Barron (eds.), *Scientific Creativity: Its Recognition and Development,* pp. 44–52. New York: Wiley.

Harry, J., and Goldner, N. S. (1972). The null relationship between teaching and research. *Sociology of Education* 45(1): 47–60.

Hayes, J. R. (1971). Research, teaching, and faculty fate. *Science* 172: 227–230.

Hicks, R. A. (1974). The relationship between publishing and teaching effectiveness. *California Journal of Educational Research* 25(3): 140–146.

Hoffman, R. A. (1984a). An assessment of the teaching, research, and service function of a college faculty. *Journal of Research and Development in Education* 17(2): 51–54.

Hoffman, R. A. (1984b). Correlates of faculty performance. *College Student Journal* 18(1): 164–168.

Hoyt, D. P. (1974). Interrelationships among instructional effectiveness, publication record, and monetary reward. *Research in Higher Education* 2(1): 81–88.

Hoyt, D. P., and Spangler, R. K. (1976). Faculty research involvement and instructional outcomes. *Research in Higher Education* 4(2): 113–122.

Hyman, H. (1955). *Survey Design and Analysis.* Glencoe, Ill.: The Free Press.

Jauch, L. R. (1976). Relationships of research and teaching: Implications for faculty evaluation. *Research in Higher Education* 5(1): 1–13.

Jencks, C., and Riesman, D. (1968). *The Academic Revolution.* Garden City, N.Y.: Doubleday.

Kulik, J. A., and Kulik, C. C. (1974). Student ratings of instruction. *Teaching of Psychology* 1(2): 51–57.

Lasher, H., and Vogt, K. (1974). Student evaluation: Myths and realities. *Improving College and University Teaching* 22(4): 267–269.

Leary, L. (1959). The scholar as teacher. *School and Society* 87: 362–363.

Light, R. J., and Pillemer, D. B. (1984). *Summing Up: The Science of Reviewing Research.* Cambridge, Mass.: Harvard University Press.

Linsky, A. S., and Straus, M. (1975). Student evaluation, research productivity, and eminence of college faculty. *Journal of Higher Education* 46(1): 89–102.

Love, J. A., Minnis, D. L. Sandoval, J. H., and Yonge, G. D. (1977). An instrument to describe college teaching: A reliability study. Paper read at the annual meeting of the American Educational Research Association.

Love, J. A., Sandoval, J., and Cohen, E. (1978). The generalizability and concurrent validity of an instrument to describe college teaching. Paper read at the annual meeting of the American Educational Research Association.

Marquardt, R. A., McGann, A. F., and Jakubauskas, E. B. (1975). Academic clients, scholarly contributions, faculty compensation and performance criteria or "Shouldn't We Take One More Look Before We Leap . . . ?" *AACSB Bulletin* 12(1): 13–17.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology* 76(5): 707–754.

Marsh, H. W., and Overall J. W. (1979). Validity of students' evaluations of teaching: A comparison with instructor self evaluations by teaching assistants, undergraduate faculty and graduate faculty. Paper read at the annual meeting of the American Educational Research Association. (ERIC Document Reproduction Service No. ED 177 205)

Maslow, A. H., and Zimmerman, W. (1956). College teaching ability, scholarly activity and personality. *Journal of Educational Psychology* 47(3): 185–189.

McCullagh, R. D., and Roy, M. R. (1975). The contribution of noninstructional activities to college classroom teacher effectiveness. *Journal of Experimental Education* 44(1): 61–70.

McDaniel, E. D., and Feldhusen, J. F. (1970). Relationships between faculty ratings and indexes of service and scholarship. *Proceedings of the 78th Annual Convention of the American Psychological Association* 5: 619–620.

McGrath, E. J. (1962). Characteristics of outstanding college teachers. *Journal of Higher Education* 33(3): 148–152.

Meltzer, L. (1956). Scientific productivity in organizational settings. *Journal of Social Issues* 12(2): 32–40.

Michalak, S. J., Jr., and Friedrich, R. J. (1981). Research productivity and teaching effectiveness at a small liberal arts college. *Journal of Higher Education* 52(6): 578–597.

Myers, C. R. (1970). Journal citations and scientific eminence in contemporary psychology. *American Psychologist* 25(11): 1041–1048.

Pezzullo, T. R., Long, J. V., and Ageloff, R. M. (1976). Paper read at the annual meeting of the Association for Institutional Research. (ERIC Document Reproduction Service No. ED 127 893)

Pillemer, D. B., and Light, R. J. (1980). Synthesizing outcomes: How to use research evidence from many studies. *Harvard Educational Review* 50(2): 176–195.

Plant, W. T., and Sawrey, J. M. (1970). Student ratings of psychology professors as teachers and the research involvement of the professors rated. *The Clinical Psychologist* 23(4): 15–16, 19.

Price, D. J. (1963). *Little Science, Big Science.* New York: Columbia University Press.

Ratz, H. C. (1975). Factors in the evaluation of instructors by students. *IEEE Transactions on Education* E-18(3): 122–127.

Riley, J. W., Jr., Ryan, B. F., and Lifshitz, M. (1950). *The Student Looks at His Teacher: An Inquiry into the Implications of Student Ratings at the College Level.* New Brunswick, N.J.: Rutgers University Press.

Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin* 85(1): 185–193.

Rosenthal, R. (1984). *Meta-Analytic Procedures for Social Research.* Beverly Hills, Calif.: Sage.

Rossman, J. E. (1976). Teaching, publication, and rewards at a liberal arts college. *Improving College and University Teaching* 24(4): 238–240.

Rushton, J. P., Murray, H. G., and Paunonen, S. V. (1983). Personality, research creativity, and teaching effectiveness in university professors. *Scientometrics* 5(2): 93–116.

Sample, S. B. (1972). Inherent conflict between research and education. *Educational Record* 53(1): 17–22.

Schmitt, H. A. (1965). Teaching and research: Companions or adversaries. *Journal of Higher Education* 36(8): 419–427.

Schrader, W. B. (1978). Admission test scores as predictors of career achievement in psychology. GRE Board Research Report No. 76-IR. Princeton, N.J.: Educational Testing Service.

Siegfried, J. J., and White, K. J. (1973). Teaching and publishing as determinants of academic salaries. *Journal of Economic Education* 4(2): 90–99.

Smith, R., and Fiedler, F. E. (1971). The measurement of scholarly work: A critical review of the literature. *Educational Record* 22(3): 225–232.

Stallings, W. M., and Singhal, S. (1970). Some observations on the relationships between research productivity and student evaluations of courses and teaching. *American Sociologist* 5(2): 141–143.

Stallings, W. M., and Spencer, R. E. (1967). Rating of instructors in Accountancy 101 from video-type clips. Research Report No. 265. Urbana-Champaign, Ill.: Measurement and Research Division, Office of Instructional Resources, University of Illinois.

Stark, J. S. (1976). A brief review of research related to the "publish or perish" ethos. Unpublished manuscript.

Stavridis, P. G. (1972). Relationships between student ratings of teaching effectiveness and certain criteria used in promotion of college faculty. Unpublished doctoral dissertation, Arizona State University.

Stumpf, S. A., Freedman, R. D., and Aguanno, J. C. (1979). A path analysis of factors often found to be related to student ratings of teaching effectiveness. *Research in Higher Education* 11(2): 111–123.

Teague, G. V. (1981). Compatibility of teaching and research at a large land-grant university. *Improving College and University Teaching* 29(1): 33–35.

Thielens, W., Jr. (1971). The undergraduate view of professors' research: Findings from the empirical literature. Unpublished manuscript, Bureau of Applied Social Research, Columbia University.

Touq, M. (1972). The relationship between student participation in classroom discussion and student ratings of instructors at the college level. Unpublished doctoral dissertation, Purdue University.

Touq, M. S., and Feldhusen, J. F. (1974). Validity of student ratings of instructors. *College Student Journal* 8(4): 2–5.

Tracy, D. B., and Tollefson, N. (1979). A multidimensional study of effective teachers. Paper read at the annual meeting of the American Educational Research Association.

Usher, R. H. (1966). The relationship of perceptions of self, others, and the helping task to certain measures of college faculty effectiveness. Unpublished doctoral dissertation, University of Florida.

Voeks, V. W. (1962). Publications and teaching effectiveness. *Journal of Higher Education* 33(4): 212–218.

Webster, D. S. (1986). Research productivity and classroom teaching effectiveness. *Instructional Evaluation* 9(1): 14–20.

Widlak, F. W., McDaniel, E. D., and Feldhusen, J. F. (1973). Factor analysis of an instructor rating scale. Paper presented at the annual meeting of the American Educational Research Association.

Wilson, L. (1964). *The Academic Man: A Study in the Sociology of a Profession.* New York: Octagon Books.

Wood, N. J., and DeLorme, C. D., Jr. (1976). An investigation of the relationship among teaching evaluation, research and ability. *Journal of Economic Education* 7(2): 77–80.

Wood, P. H. (1978). Student and peer ratings of college teaching and peer ratings of research and service: Four years of departmental evaluation. Paper read at the annual meeting of the American Educational Research Association.

Woodburne, L. S. (1952). The qualifications of superior faculty members. *Journal of Higher Education* 23(7): 377–382.

Received February 3, 1987

APPENDIX

This appendix presents results of studies that have data comparing the research and scholarly productivity of faculty members with the evaluations they received from students on *specific* instructional aspects of their teaching and the courses they taught (rather than on *overall* effectiveness of the teachers and courses). Measures used in each study are given in an abbreviated format. Descriptions of the samples used or populations studied can be found in Table 1 in the text.

An asterisk in front of the citation to a study indicates that the relationship between research productivity and the specific evaluation either was originally given as a product-moment correlation (r) in the study or that the statistic(s) presented in the study for the relationship could be converted into an r by procedures suggested in Glass, McGaw, and Smith (1981, see especially Table 5.8, pp. 149–150). The Z (standard normal deviate) associated with each of these r's has been calculated using procedures suggested by Rosenthal (1984, see especially pp. 106–107).

These r's and Z's are the components of the average r's and combined Z's given in Table 2 in the text. Note that results are given in this appendix for those few cases (indicated by omission of asterisks) where an r was not reported in the study and where one could not be estimated from the data that were presented. These findings are not incorporated into Table 2, but have been included in the appendix for purposes of comprehensiveness.

The instructional dimensions into which the results from the studies were classified are the same ones used in two different reviews by Feldman (1983, 1984), as based on a still earlier analysis (Feldman, 1976b). Multi-item scales—always called "scales" to distinguish them from single evaluation items—are coded into the instructional dimensions by considering the content of their individual items rather than by the names that have been given to them by the researcher(s). Because some of the items or scales measuring specific evaluations fit into more than one of the instructional dimensions, the associations between them and research productivity are coded in more than one dimension in this appendix, as the entries make clear. In order not to "overcount" these associations when averaging the correlations within a dimension, each association has a weight corresponding to the inverse of the number of dimensions for which it is relevant (for examples, a weight of $\frac{1}{3}$ if the same association appears in 3 different instructional dimensions, $\frac{1}{2}$ if 2 dimension, and 1 if only 1 dimension). These weights, shown for each asterisked entry, were used when constructing Table 2.

RELATIONSHIP BETWEEN RESEARCH OR SCHOLARLY PRODUCTIVITY OF FACULTY MEMBERS AND STUDENTS' EVALUATIONS ON SPECIFIC INSTRUCTIONAL DIMENSIONS

Instructional Dimension No. 1: Teacher's Stimulation of Interest in the Course and Its Subject Matter

(*) Aleamoni and Yimer (1973): Publication × "Interest and Attention" Scale: average $r = -.04$; $Z = -.204$. (Weight: 1)

Cornwell (1974): "Research activity" × stimulated interest: point-biserial correlation = .16 and $Z = 1.602$ based on F; but direction of association cannot be determined.

(*) Freedman, Stumpf, and Aguanno (1979): Publication × "Instructor in Class" Scale (adjusted for instructor's grading policy): $r = +.21$; $Z = +2.393$. Also coded in Dimensions No. 6 and No. 17. (Weight: 1/3)

(*) Friedrich and Michalak (1983): Dean's and chairperson's evaluation of research ability × interestingness of presentation: $r = +.10$; $Z = +0.851$. (Weight: 1)

(*) Hayes (1971): Publication × presentation (or class) stimulating. The author writes that no significant relationship was found, but exact data are not presented (preventing calculation of r); r considered to be .00 and Z considered to be 0.000. (Weight: 1)

(*) Marsh and Overall (1979): Self-rating on "scholarly production" × "Instructor Enthusiasm" Scale: $r = +.02$; $Z = +0.269$. Also coded in Dimension 2. (Weight: 1/2)

(*) McDaniel and Feldhusen (1970): Publication × stimulation: average $r = +.05$; $Z = +0.430$. (Weight: 1)

(*) Rossman (1976): Publication × most stimulating course taken at the university: $r = +.19$; $Z = +2.100$. (Weight: 1)

(*) Stavridis (1972): Publication × arouses interest in the subject: $r = +.17$; $Z = +0.938$. (Weight: 1)

Instructional Dimension No. 2: Teacher's Enthusiasm (for Subject or for Teaching)

(*) Friedrich and Michalak (1983): Dean's and chairperson's evaluation of research ability × instructor's enthusiasm: $r = +.04$; $Z = +0.340$. (Weight: 1)

(*) Harry and Goldner (1976): Publication × enthusiastic: $r = +.19$; $Z = +1.682$.

(*) Linsky and Straus (1975): Publication and citations × instructor's interest in subject: average (weighted) $r = +.07$; average (weighted) $Z = +0.580$. (Weight: 1)

(*) Marsh and Overall (1979): Self-rating on "scholarly production" × "Instructor Enthusiasm" Scale: $r = +.02$; $Z = +0.269$. Also coded in Dimension No. 1. (Weight: 1/2)

Instructional Dimension No. 3: Teacher's Knowledge of Subject

(*) Friedrich and Michalak (1983): Dean's and chairperson's evaluation of research ability × knowledgeability of instructor: $r = -.26$; $Z = -2.243$. (Weight: 1)

(*) Linsky and Straus (1975): Publication and citations × instructor's knowledge: average (weighted) $r = +.20$; average (weighted) $Z = +3.393$. (Weight: 1)
(*) McDaniel and Feldhusen (1970): Publication × knowledge of subject matter: $r = +.11$; $Z = +0.949$. (Weight: 1)
(*) Riley, Ryan, and Lifshitz (1950): Publication × knowledge of subject: r (tetrachoric) = $+.46$; $Z = +9.421$. (Weight: 1)
(*) Stavridis (1972): Publication × how well teacher knows the subject: $r = +.56$; $Z = +3.279$. (Weight: 1)

Instructional Dimension No. 4: Teacher's Intellectual Expansiveness (and Intelligence)

(*) Marsh and Overall (1979): Self-rating on "scholarly production" × "Breadth of Coverage" Scale: $r = +.21$; $Z = +2.856$. (Weight: 1)
(*) Stavridis (1972): Publication × how well teacher relates the work of the course with other areas of knowledge: $r = +.08$; $Z = +0.439$. (Weight: 1)

Instructional Dimension No. 5: Teacher's Preparation; Organization of the Course

(*) Centra (1983, Study 1): Publication × "Course Organization and Planning" Scale: average (weighted) $r = +.05$; $Z = +2.725$. Also coded in Dimension No. 9. (Weight: 1/2)
Cornwell (1974): "Research activity" × well prepared: point-biserial $r = .00$; $Z = 0.000$. Note: Although these results are potentially includable, they are not used in constructing Table 2 in the text; the associations for the other specific evaluations in the study cannot be included because their directions are unknown—see the entries for Cornwell (1974) in Dimensions No. 1 and No. 8—and so the results for this dimension are also excluded.
(*) Frey (1976): Citations × "Pedagogical Skill" Factor Scale: $r = +.37$; $Z = +2.419$. Also coded in Dimensions No. 6 and No. 12. (Weight: 1/3)
(*) Friedrich and Michalak (1983): Dean's and chairperson's evaluation of research ability × degree of preparation: $r = +.20$; $Z = +1.714$. (Weight: 1)
(*) Marsh and Overall (1979): Self-rating on "scholarly production" × "Organization" Scale: $r = +.18$; $Z = +2.441$. Also coded in Dimension No. 6. (Weight: 1/2)
(*) McDaniel and Feldhusen (1970): Publication × organization: average $r = +.11$; $Z = +0.949$. (Weight: 1)
(*) Riley, Ryan, and Lifshitz (1950): Publication × organization of subject matter: r (tetrachoric) = $+.22$; $Z = +4.340$. (Weight: 1)
(*) Stavridis (1972). Publication × how well prepared for class meetings: $r = +.23$; $Z = +1.276$. (Weight: 1)

Instructional Dimension No. 6: Clarity and Understandableness

(*) Aleamoni and Yimer (1973): Publication × "Course Content" Scale: average $r = +.08$; $Z = +0.424$. Also coded in Dimension No. 10. (Weight: 1/2)

(*) Centra (1983, Study 1). Publication × "Communication" Scale: average (weighted) $r = +.03$; $Z = +1.634$. Also coded as Dimension No. 17. (Weight: 1/2)

(*) Freedman, Stumpf, and Aguanno (1979). Publication × "Instructor in Class" Scale (adjusted for instructor's grading policy): $r = +.21$; $Z = +2.393$. Also coded as Dimensions No. 1 and No. 17. (Weight: 1/3)

(*) Frey (1976): Citations × "Pedagogical Skill" Factor Scale: $r = +.37$; $Z = +2.419$. Also coded in Dimensions No. 5 and No. 12. (Weight: 1/3)

(*) Friedrich and Michalak (1983): Dean's and chairperson's evaluation of research ability × clarity of presentation: $r = +.19$; $Z = +1.627$. (Weight: 1)

(*) Harry and Goldner (1982): Publication × instructor is clear: $r = +.10$; $Z = +0.874$. (Weight: 1)

(*) Linsky and Straus (1975): Publication and citations × course coherence: average (weighted) $r = -.02$; average (weighted) $Z = -0.191$. (Weight: 1)

(*) Marsh and Overall (1979): Self-rating on "scholarly production" × "Organization" Scale: $r = +.18$; $Z = +2.441$. Also coded as Dimension No. 5. (Weight: 1/2)

(*) McDaniel and Feldhusen (1970): Publication × presentation of course material: $r = +.04$; $Z = +0.344$. (Weight: 1)

(*) Riley, Ryan, and Lifshitz (1950): Publication × ability to explain: r (tetrachoric) = $+.14$; $Z = +2.742$. (Weight: 1)

Instructional Dimension No. 7: Teacher's Elocutionary Skills

No entries.

Instructional Dimension No. 8: Teacher's Sensitivity to, and Concern with, Class Level and Progress

(*) Centra (1983, Study 1): Publication × "Faculty-Student Interaction" Scale: average (weighted) $r = +.06$; $Z = +3.271$. Also coded in Dimensions No. 16, No. 17, and No. 19. (Weight: 1/4)

Cornwell (1974): "Research activity" × interested in students' progress: point-biserial correlation = .12 and $Z = 1.180$ based on F; but direction of association cannot be determined.

(*) Stavridis (1972): Publication × how well teacher adjusts to students' level of competence: $r = +.07$; $Z = +0.384$. (Weight: 1)

Instructional Dimension No. 9: Clarity of Course Objectives and Requirements

(*) Centra (1983, Study 1): Publication × "Course Organization and Planning" Scale: average (weighted) $r = +.05$; $Z = +2.725$. Also coded in Dimension No. 5. (Weight: 1/2)

(*) Friedrich and Michalak (1983): Dean's and chairperson's evaluation of research ability × explicitness of requirements: $r = +.27$; $Z = +2.333$. (Weight: 1)

(*) McDaniel and Feldhusen (1970): Publication × policies: average $r = +.06$; $Z = +0.517$. (Weight: 1)

(*) Stavridis (1972): Publication × how well objectives of the course were explained: $r = +.27$; $Z = +1.506$. (Weight: 1)

Instructional Dimension No. 10: Nature and Value of the Course Material (Including Its Usefulness and Relevance)

(*) Aleamoni and Yimer (1973): Publication × "Course Content" Scale: average $r = +.08$; $Z = +.409$. Also coded in Dimension No. 6. (Weight: 1/2)

(*) Bausell and Magoon (1972): Publication and grants × text quality and course relevance: all relevant r's $< .10$, counted as average $r = .00$; $Z = 0.000$. (Weight: 1)

(*) Centra (1983, Study 1): Publication × "Textbooks and Reading" Scale: average $r = +.02$; $Z = +1.089$. Also coded in Dimension 11. (Weight: 1/2)

(*) Linsky and Straus (1975): Publication and citations × course content and value of readings: average (weighted) $r = +.06$; average (weighted) $Z = +1.320$. (Weight: 1)

(*) Marsh and Overall (1979): Self-rating on "scholarly production" × "Assignments" Scale: $r = +.17$; $Z = +2.304$. Also coded in Dimension No. 11. (Weight: 1/2)

(*) McDaniel and Feldhusen (1970): Publication × text and applications: average $r = +.09$; $Z = +0.776$. (Weight: 1)

Instructional Dimension No. 11: Nature and Usefulness of Supplementary Materials and Teaching Aids

(*) Centra (1983, Study 1): Publications × "Textbooks and Reading" Scale: average $r = +.02$; $z = +1.089$. Also coded in Dimension No. 10. (Weight: 1/2)

(*) Linsky and Straus (1975): Publication and citations × value of papers: average (weighted) $r = +.07$; average (weighted) $Z = +1.262$. (Weight: 1)

(*) Marsh and Overall (1979): Self-rating on "scholarly production" × "Assignments" Scale: $r = +.17$; $Z = +2.304$. Also coded in Dimension No. 10. (Weight: 1/2)

Instructional Dimension No. 12: Perceived Outcome or Impact of Instruction

(*) Frey (1976): Citations × "Pedagogical Skill" Factor Scale: $r = +.37$; $Z = +2.419$. Also coded in Dimensions No. 5 and No. 6. (Weight: 1/3)

(*) Friedrich and Michalak (1983): Dean's and chairperson's evaluation of research ability × students' perceived acquisition of knowledge, perceived development of rational and critical skills, and perceived development of intellectual independence: average $r = +.13$; $Z = +1.108$. (Weight: 1)

(*) Harry and Goldner (1972): Publication × increased interest in course matter: $r = +.05$; $Z = +0.439$. (Weight: 1)

(*) Hoyt and Spangler (1976): Department head's rating of research involvement (time and accomplishment) × students' perceived progress on ten different objectives plus increase in positive feelings toward field of study represented by course (controlled for students' motivation to take the course): average (weighted) r (eta) = +.03; average (weighted) Z based on $F = +0.223$. (Weight: 1)

(*) Marsh and Overall (1976): Self-rating on "scholarly production" × "Learning/Value" Scale: $r = +.12$; $Z = +1.620$. Also coded in Dimension No. 17. (Weight: 1/2)

(*) Rossman (1976): Publication × faculty member contributing most to student's emotional and/or personal development: $r = +.14$; $Z = +1.541$. (Weight: 1)

Instructional Dimension No. 13: Instructor's Fairness; Impartiality of Evaluation of Students; Quality of Examinations

(*) Centra (1983, Study 1): Publication × "Tests and Exams" Scale: average (weighted) $r = +.04$; $Z = +2.179$. (Weight: 1)

(*) Freedman, Stumpf, and Aguanno (1979): Publication × "Graded Assignments/Examinations" Scale (adjusted for instructor's grading policy): $r = -.04$; $Z = -0.451$. (Weight: 1)

(*) Frey (1976): Citations × "Rapport" Factor Scale: $r = -.23$; $Z = -1.474$. Also coded in Dimensions No. 16 and No. 19. (Weight: 1/3)

(*) Linsky and Straus (1975); Publication and citations × exam quality and fairness: average (weighted) $r = -.02$; average (weighted) $Z = -0.494$. (Weight: 1)

(*) Marsh and Overall (1979): Self-rating on "scholarly production" × "Examination" Scale: $r = +.04$; $Z = +0.538$. Also coded in Dimension No. 15. (Weight: 1/2)

(*) McDaniel and Feldhusen (1970): Publication × grading and exams: average $r = +.07$; $Z = +0.603$. (Weight: 1)

Instructional Dimension No. 14: Personality Characteristics ("Personality") of the Teacher

(*) Friedrich and Michalak (1983): Dean's and chairperson's evaluation of research ability × instructor's personality: $r = +.12$; $Z = +1.022$. (Weight: 1)

Instructional Dimension No. 15: Nature, Quality, and Frequency of Feedback from the Teacher to Students

(*) Friedrich and Michalak (1983): Dean's and chairperson's evaluation of research ability × feedback on performance and promptness in returning work: average $r = +.08$; $Z = +0.680$. (Weight: 1)

(*) Marsh and Overall (1979): Self-rating on "scholarly production" × "Examinations" Scale: $r = +.04$; $Z = +0.538$. Also coded in Dimension No. 13. (Weight: 1/2)

Instructional Dimension No. 16: Teacher's Encouragement of Questions and Discussion, and Openness to Opinions

(*) Centra (1983, Study 1): Publication × "Faculty-Student Interaction" Scale: average (weighted) $r = +.06$; $Z = +3.271$. Also coded in Dimensions No. 8, No. 17, and No. 19. (Weight: 1/4)

(*) Frey (1976): Citations × "Rapport" Factor Scale: $r = -.23$; $Z = -1.474$. Also coded in Dimensions No. 13 and No. 19. (Weight: 1/3)

(*) Marsh and Overall (1979): Self-rating on "scholarly production" × "Group Interaction" Scale: $r = +.04$; $Z = +0.538$. (Weight: 1)

(*) McDaniel and Feldhusen (1970): Publications × chance to question and discussions: average $r = +.04$; $Z = +0.344$. (Weight: 1)

(*) Stavridis (1972): Publications × how tolerant teacher is of opinions of others: $r = -.02$; $Z = -0.110$. (Weight: 1)

Instructional Dimension No. 17: Intellectual Challenge and Encouragement of Independent Thought (by the Teacher and the Course)

(*) Centra (1983, Study 1): Publication × "Communication" Scale: average (weighted) $r = +.03$; $Z = +1.634$. Also coded in Dimension No. 6. (Weight: 1/2)

Publication × "Faculty-Student Interaction" Scale: average (weighted) $r = +.06$; $Z = +3.271$. Also coded in Dimensions No. 8, No. 16, and No. 19. (Weight: 1/4)

(*) Freedman, Stumpf, and Aguanno (1979): Publication × "Instructor in Class" Scale (adjusted for instructor's grading policy): $r = +.21$; $Z = +2.393$. Also coded in Dimensions No. 1 and No. 6. (Weight: 1/3)

(*) Friedrich and Michalak (1983): Dean's and chairperson's evaluation of research ability × intellectual challenge: $r = +.08$; $Z = +0.670$. (Weight: 1)

(*) Marsh and Overall (1979): Self-rating on "scholarly production" × "Learning/Value" Scale: $r = +.12$; $Z = +1.620$. Also coded in Dimension No. 12. (Weight: 1/2)

Instructional Dimension No. 18: Teacher's Concern and Respect for Students; Friendliness of the Teacher

(*) Freedman, Stumpf, and Aguanno (1979): Publication × "Instructor in General" Scale (adjusted for instructor's grading policy): $r = +.18$; $Z = +2.045$. Also coded in Dimension No. 19 (Weight: 1/2)

(*) Linsky and Straus (1975): Publication and citations × instructor's personalization (degree to which instructor seems personally responsive to students): average $r = -.06$; $Z = -1.640$. (Weight: 1)

(*) Marsh and Overall (1979): Self-rating on "scholarly production" × "Individual Rapport" Scale: $r = +.06$; $Z = +0.808$. Also coded in Dimension No. 19. (Weight: 1/2)

(*) *McDaniel and Feldhusen (1970):* Publication × rapport: average $r = +.01$; $Z = +0.086$. (Weight: 1)

Riley, Ryan, and Lifshitz (1950): Publication × attitude toward students: relationship is reported to be inverse, but exact data are not given.

(*) *Stavridis (1972):* Publication × feeling between instructor and students: $r = +.13$; $Z = +0.738$. (Weight: 1)

Instructional Dimension No. 19: Teacher's Availability and Helpfulness

(*) *Centra (1983, Study 1):* Publication × "Faculty-Student Interaction" Scale: average (weighted) $r = +.06$; $Z = +3.271$. Also coded in Dimensions No. 8, No. 16, and No. 17. (Weight: 1/4)

(*) *Freedman, Stumpf, and Aguanno (1979):* Publication × "Instructor in General" Scale (adjusted for instructor's grading policy): $r = +.18$; $Z = +2.045$. Also coded in Dimension No. 18. (Weight: 1/2)

(*) *Frey (1976):* Citations × "Rapport" Factor Scale: $r = -.23$; $Z = -1.474$. Also coded in Dimensions No. 13 and No. 16. (Weight: 1/3)

(*) *Friedrich and Michalak (1983):* Dean's and chairperson's evaluation of research ability × availability outside class: $r = +.07$; $Z = +0.595$. (Weight: 1)

(*) *Marsh and Overall (1979):* Self-rating on "scholarly production" × "Individual Rapport" Scale: $r = +.06$; $Z = +0.808$. Also coded in Dimension No. 18. (Weight: 1/2)

(*) *McDaniel and Feldhusen (1970):* Publication × outside help: average $r = +.04$; $Z = +0.344$. (Weight: 1)

(*) *Stavridis (1972):* Publication × how students feel they are able to get personal help in the course: $r = -.17$; $Z = -0.938$. (Weight: 1)