DOCUMENT RESUME

ED 379 705 CS 508 824

AUTHOR Allen, Mike
TITLE Research Productivity and Positive Teaching Evaluations: Examining the Relationship Using Meta-Analysis.
PUB DATE Feb 95
NOTE 46p.; Paper presented at the Annual Meeting of the Western States Communication Association (Portland, OR, February 10-14, 1995).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143) -- Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *College Faculty; Departments; Higher Education; Meta Analysis; *Productivity; *Publish or Perish Issue; *Research; *Speech Communication; Student Attitudes; Student Evaluation of Teacher Performance; *Teacher Effectiveness
IDENTIFIERS Educational Issues

ABSTRACT

Arguments about trade-offs existing between teaching and research affect much of the communication discipline as scholars engage in arguments about the future directions of departments. A study summarized more than 40 quantitative studies and found a small heterogeneous positive correlation between teaching effectiveness and research productivity. Positive teaching evaluations correlate with increased research productivity. While the finding should not be interpreted as direct evidence of any causality between the variables, the evidence points to an association that deserves consideration. While the correlation is small, the association remains positive, suggesting that research productivity does not necessarily contradict efforts at quality teaching. The finding warrants a more thorough understanding of those features associated with both increased research productivity and positive teaching evaluations. Building a strong and effective communication department requires that the philosophical underpinnings take into account the requirements and potential trade-offs between research productivity and teaching effectiveness. Contains 4 notes, 120 references, and 3 tables of data.) (RS)

RESEARCH PRODUCTIVITY AND POSITIVE TEACHING EVALUATIONS: EXAMINING THE RELATIONSHIP USING META-ANALYSIS

by

Mike Allen
Department of Communication
University of Wisconsin-Milwaukee
Milwaukee, WI 53201

Manuscript presented at the
Western Communication Association Convention

Portland, Oregon

February, 1995

ABSTRACT

RESEARCH PRODUCTIVITY AND POSITIVE TEACHING EVALUATIONS:

EXAMINING THE RELATIONSHIP USING META-ANALYSIS

A long-term controversy surrounds whether college faculty face a trade-off between producing research and offering quality instruction. The debate assumes that a college professor cannot combine excellence in teaching and research. This summary of more than 40 quantitative studies found a small heterogeneous positive correlation between teaching effectiveness and research productivity. Positive teaching evaluations correlate with increased research productivity. While the finding should not be interpreted as direct evidence of any causality between the variables, the evidence points to an association that deserves consideration. While the correlation is small, the association remains positive, suggesting that research productivity does not necessarily contradict efforts at quality teaching. The finding warrants a more thorough understanding of those features associated with both increased research productivity and positive teaching evaluations.

Almost all universities evaluate faculty based on three standard areas: (a) research, (b) teaching, and (c) service. The trend in recent years seems to favor increasing the prominence and importance of research as a standard for tenure and merit for college professors (the publish or perish syndrome). A long-standing controversy involves whether trade-offs exist between research and teaching for college faculty (Faia, 1976; Grant, 1971; Hammond, Meyer, & Miller, 1969; Harry & Goldner, 1972; Jauch, 1976; Kurland, 1961; Lavis, 1992; Martin & Berry, 1961; Rodden, 1993; Schachter, 1991; Smith, 1961; Woodburne, 1952). As the institutional emphasis on research appears to increase, the controversy surrounding whether such increased expectations for faculty research productivity diminished the quality of teaching becomes more important.

The reasons for increasing the emphasis on research are probably several and varied depending on the various internal pressures of the institution. The increased desire for outside funding places an emphasis on those attributes believed related to the potential for generating funds (research capability) (Burgoon, 1988). One communication scholar (Burgoon, 1989) argues for a separation of those departments interested in research from those departments devoted to instruction (the place of service within this debate seems lost). If communication departments emphasize research and external funding this may be deleterious to the quality of instruction provided by those departments.

The belief or search for an objective merit system for promotion, tenure, and salary tends to emphasize those things objectively quantifiable (number of publications) to those items perceived more difficult to objectify (quality of instruction). The desire by administrations and faculty for higher "status" leads to an emphasis on research in a belief that more research productivity improves status within the academic community (for example several scholars in communication have commented on the need for scholars in the discipline to produce "research books"). Several efforts to quantify the contribution of communication scholars to the discipline exists that notes where prolific scholars obtained their degrees and the institutions currently employing these persons (Barker, Hall, Roach, & Underberg, 1979; 1980; 1981; Barker, Ray, Watson, & Hall, 1988; Hickson, 1990; Hickson, Stacks, & Amsbary, 1989; 1992; 1993; Stacks & Hickson, 1983; Watson, Barker, Ray, & Hall, 1988). This type of scoreboard system has not been without criticism (Brown, Blair, & Baxter, 1004; Erickson, Fleuriet, & Hosman, 1993), but the ability to provide a metric of comparison for research output leads to a sense of direct comparison of the value of various scholars.

The ability to create "status" for the quality of instruction seems difficult for the purposes of direct institutional comparisons. Identifying the quality of instruction becomes viewed by individual faculty as a political football within departments. Instructors become concerned that factors affecting the favorableness of student evaluations include: (a) the grade point average of the class, (b) the level of the class, (c) the size of the class, and (d) the content of the class. Without considering a host of potential moderating factors the ability to compare directly evaluations seems difficult. Some faculty the author has known refer to student evaluations of teaching as the "Nielson ratings" of academe. This perspective when combined with the lack of public availability of records makes comparing departments or individual faculty from different institutions difficult, if not impossible. The purpose of this review is not to offer a direct critique or support for the validity of any methods of instructional evaluation.[1] This review only intends to use existing data as a basis for comparison.

The arguments about trade-offs existing between teaching and research affect much of the discipline of communication as scholars engage in arguments about the future directions of departments. One argument concludes that increased status for communication departments come from the ability to generate money for research. Some scholars argue that the inability of communication researchers to generate valuable theory and effect the intellectual climate of other disciplines indicates a weakness of the field. Other academicians suggest the ability of communication departments to attract quality students and provide them the skills for employment and empowerment create a strong justification for communication departments. These arguments either directly or indirectly consider the connection between quality of instruction and research productivity. The concern taken with students and/or research, especially when evaluating the efforts of our faculty, require consideration of the relationship between research productivity and teaching effectiveness. Building a strong and effective communication department requires that the philosophical underpinnings consider the requirements and potential trade-offs between research productivity and teaching effectiveness. The next two sections explore arguments about the nature of the connection.

ARGUMENTS THAT A RESEARCH EMPHASIS UNDERMINES POSITIVE TEACHING EVALUATIONS

One issue advanced by those persons critical of requiring a great deal of research by faculty is that the trend away from the classroom and into research changes the role of college faculty. The argument runs that time spent on research and writing becomes subtracted from the time spent with students in classes, laboratory sessions, and office hours. Much of the research arguably could never benefit the students (or society) and therefore the emphasis on research becomes misplaced. The research fails to improve classroom performance or the mind of the instructor. The research serves to increase the economic value and/or professional ego of the faculty member. Sykes, in his book Profscam (1988) argues particularly against communication scholarship (see pp. 111-113 for his analysis of articles appearing in Human Communication Research and Southern Speech Communication Journal). His argument becomes echoed by Bloom's (1987) argument about the tendency of research to benefit faculty members and no one else. This argument against mindless publication receives support from persons in communication (Erickson, Fleuriet, & Hosman, 1993) when they point out as a myth that "research is more rewarding than teaching" (p. 335). When the pressures of producing research become an end, without true intellectual advancement, the expenditure of time in the endeavor may reduce time spent on other duties.

While the faculty member conducting research enriches one's status in the university (among certain circles), the result impoverishes the student in the classroom. The time devoted to the laboratory, word processor, conventions, editing, and/or library become subtracted from the time spent on lecture notes, advising students, and improving methods of evaluation (tests, papers, quizzes). The publication becomes an end, rather than the idea the published article should contain. Scholars talk about the LPU (Least Publishable Unit), the smallest amount of information that justifies a separate article (another listing on the annual report of activities and a longer VITA). The image of "publication monsters" exists, generating paper trails into the nothingness of larger egos, benefiting no undergraduate (sometimes not even graduate) students, only the faculty member.

This argument runs that there exists a zero sum resource called time with trade-offs existing between time spent on teaching and research. Most research information cannot provide insight to help students in the classroom or improve teaching. The professor creates a set of lecture notes, chooses a textbook, creates a standardized midterm and final, and simple paper assignments. The instructor creates a course that receives little updating, new thought, or originality. The textbook companies provide instructor manuals with exercises, overheads, videocassettes, tests, quizzes, discussion questions, all the material necessary for instruction. The teacher need not think, but only act to convey material already supplied and organized for transmission and consumption. The teaching languishes (by remaining stagnant) while the research publication process continues.

Another issue involves the separation that a research emphasis creates between the student and the instructor. Office hours, advising, and student activities require time from the instructor. A professor working on student-oriented activities (director of undergraduate studies, forensics director, mock trial coordinator, adviser to a student-run paper or radio station, or other professional student organizations like Women in Communication Incorporated) takes time away from research. Research professors whose reputation, salary, and job security are not dependent on these activities could view working with students as a waste of time. Students become seen as lab rats for experiments, research assistants, or some other type of research aid. The professor views normal teaching as onerous and nonclassroom contact as undesirable. The very framework of the institutional pressures contributes to a deteriorating relationship between faculty and students.

The previously described process reduces scores on teaching evaluations. As faculty spend more time away from the classroom and students, the quality of instruction languishes and erodes. Students become alienated from the faculty as the faculty become inaccessible and unavailable to students. The key to this line of argument is that faculty members possess little incentive to provide quality instruction and that without such an incentive the rewards of research become alluring. Since there exists no limit to the amount of research the professor can produce, and the more the better, the efforts of the scholar become heavily tilted favoring research (the creation of a giant paper chase).

ARGUMENTS THAT A RESEARCH EMPHASIS IMPROVES TEACHING EVALUATIONS

The alternative line of argument suggests a connection between research productivity and teaching quality. Teaching many topics within the university, especially the field of communication, requires sophisticated knowledge of a constantly changing subject matter. A top researcher, arguably, provides better, up-to-date, and accurate information than someone not involved in research. This occurs because the researcher remains active and involved on the cutting edge of theory and knowledge while the nonresearcher does not. This argument assumes a relationship between teaching and research permitting the researcher to bring the benefits of the research into the classroom. Research shows not only knowledge but shows a dedication to the content and material not possible for a person that only teaches.

The distinction exists between what might be considered the elementary and secondary versus college models for instruction. The argument is that college instructors must offer more than simply quality instruction, the actual content of the information changes and the college instructor should contribute to that process of change. No one expects a high school or elementary teacher to generate knowledge claims, the emphasis is placed on quality of instruction.

Consider the nature of teaching some content area at the college level. An instructor selects a textbook, writes lecture notes, tests, paper assignments, and relevant supporting documents for students. After three years, the instructor has a course to the point they feel is complete, rigorous, and intellectually stimulating. But theories and knowledge changes as more becomes known, considered, and reconsidered. After a few years the need to change particular features of the course is necessary. Consider a recent examination of textbooks using meta-analyses (Allen & Preiss, 1990) to evaluate their accuracy. That article in part required communication textbook authors to examine research more closely and a reexamination two later editions of those textbooks (Osborn & Osborn, 1994; Ross, 1992) provides examples of changes in the treatment of particular research literature, improving the accuracy of textbooks.

Specialization involved in research permits greater depth of knowledge for students to capitalize on. The ability of instructors to act as resource persons depends on the ability to grasp current material. The faculty member involved in research benefits the student and the community by providing such material. The enthusiasm necessary for research should spill over into the classroom as the researcher provides a commitment and a stimulation not possible if they were not engaged in such research.

EXAMINING THIS ISSUE EMPIRICALLY

A host of quantitative examinations on this issue permits a synthesis using meta-analysis. Because a large number of investigations exists on this issue should not be surprising since the data to examine the relationship is routinely collected. Teaching evaluations (student, peer, and administrative) occur with regularity and much effort involved in gathering and accumulating such information. The normal process of merit, tenure, and promotion require some effort at the documentation of instructional quality. Research productivity generally receives evaluation within the faculty merit system, often yearly. The system generates essentially a regular system of records permitting a researcher to compare the various methods of rating research productivity. The basis typically takes some form of quantitative measure of the amount of research (either in grant dollars or number of publications) produced by a scholar.

A previous meta-analysis does exist on this topic (Feldman, 1987). However, there exists many justifications for both replicating and extending the earlier analysis. First, no literature search method details exist in the manuscript, making it impossible to know to what extent the author searched the relevant literature and under what conditions. There is no way of knowing the procedures of the literature search, what definitions used for the process, nor the success of the literature search. A literature search should be explicit to permit evaluation of the methods, completeness, and adequacy (Preiss & Allen, 1994).

Second, the report provides no homogeneity test for the average correlation produced was conducted. This lack of testing constitutes a serious weakness because the ability to interpret any results of a meta-analysis depends on a negative outcome for the test of moderator variables. Feldman (1987) used an early form the technique developed by Glass, McGaw, and Smith (1981) that did not provide a test for moderator variables. The interpretation of the mean effect size in any meta-analysis depends on the homogeneity of the mean correlation (Dillard & Hale, 1993; Hunter & Schmidt, 1991). The interpretation of an average correlation based on a heterogeneous set of correlations would be like interpreting a main effect in an ANOVA in the presence of an interaction (Hunter, Hamilton, & Allen, 1989). The interpretation of a main effect under those conditions remains problematic (Winer, 1971). The failure to perform a test indicates that no information is known about the possible existence of the moderator variables and any conclusion remains limited.

Third, Feldman's (1987) treatment of data reports with incomplete information was to exclude the data set. If a manuscript reported a nonsignificant finding Feldman did not report any effect for the data. The procedure creates a potential upward bias in the average correlation because it systematically excluded data with small effects that tend to be nonsignificant, the resulting average correlation represents an overestimate of the true effect. Other procedures and possibilities remain to consider recovering and including data within the meta-analysis literature (Allen, Hunter, & Donohue, 1989; Boster & Mongeau, 1984).

Finally, all research requires replication. Even a meta-analysis requires replication before accepting the results as definitive or authoritative (Allen & Preiss, 1993). If any error exists by the person conducting the analysis, the replication should reveal the error and permit an assessment of it. Such replications occur frequently using meta-analysis on a variety of topics: (a) persuasiveness of fear appeals (Boster & Mongeau, 1984; Mongeau, 1994; Sutton, 1982), (b) the consistency of attitude-behavior relationships (Kim & Hunter, 1993a; 1993b; Sheppard, Hartwick, & Warshaw, 1988), (c) the impact of foot-in-the-door or door-in-the-face appeals (Beamon, Cole, Preston, Klentz, Steblay, 1988; Dillard, Hunter & Burgoon, 1984; Fern Monroe, & Avila, 1986) and the (d) persuasiveness of message sidedness (Allen, 1989; 1993; O'Keefe, 1993). The replications offer additional perspectives and reaffirmation or challenge to the findings. Scientific advancement depends on findings that demonstrate replication, even for meta-analysis.

METHOD

Literature Search

The literature search conducted used both a manual and CD ROM (Silverplatter) search of ERIC, Psychlit, and CommIndex. A manual search was conducted on the Educational Research Index. The key words used were "faculty evaluation" and "faculty promotion" as well as "productivity." All manuscript's references sections found by this search received examination for possible additional sources of information. No manuscripts existed that were authored by communication scholars, or focused particularly on communication departments. However, more than 75% of the data sets used faculty across departments at institutions, suggesting that while communication departments were not the primary focus of the investigations, communication faculty were undoubtedly included within this report. For inclusion in this meta-analysis a manuscript had to contain the following set of information:

(a) a measure of teaching effectiveness;

(b) a measure of research productivity; and

(c) statistical information permitting the calculation of an effect size, manuscripts that presented redundant data sets were not included in this analysis as separate data points.[2]

Coding of Studies

Studies were coded for relevant moderator features: (a) year of investigation, (b) method of teaching evaluation, and (c) method of measuring productivity for research.

Year

One feature of interest in this analysis is whether the year of data collection would moderate the analysis. The argument runs that the changing emphasis on research within the academic community may change the underlying relationship between research and teaching. This moderator analysis would test this assumption. The actual date of data collection and not the date of publication remains the important feature. Therefore the earliest public presentation date became the date used if the actual date of data collection was not provided. One study (Linsky & Straus, 1975) collected data over a 20 year period and became excluded from this analysis, the band of years was too wide to permit a single date to be entered for this estimate. All the coding in this section agreed with information provided by Feldman (1987) (for those manuscripts he reports).

Method of Measuring Evaluation of Teaching

The method of teaching evaluation should be considered because of the potential differences in assessing instructional effectiveness. Some studies use: (a) student evaluations, (b) peer evaluation, (c) a nomination or receipt of an award for teaching, and (d) measurement of teaching related activities. The coding of this information agreed with the representations by Feldman (1987).

Method of Measuring Research Productivity

The method of considering how productive a scholar's research output can be measured in a variety of methods. Some methods use a counting (weighted or unweighted) involving the sheer number of publications, other methods rate on a scale (1 to 5) the productivity of the person utilizing a faculty panel. Some measures consider the quality of publications as an index of productivity (using the Social Sciences Citation Index as a measure of quality by counting the number of citations of a scholar's work). One measure of productivity involved the number of grants a person had received as a faculty member. A final index takes the publications and weighs the value of the research in some manner (for example, a book is 25 point, single-authored article in national journal is 10 points, co-authored national article is 7 points, etc.,). The measurements of research coding used the following scheme of eight values: (a) number of publications, (b) grants awarded, (c) number of citations, (d) peer or chair rating of research, (e) time spent on research, (f) awards for research, (g) combination of grants and publications, and (h) the research creativity of the scholar as rated by other faculty. The coding in this section agreed completely with that information provided by Feldman (1987).

Statistical Analysis

Statistical analysis requires three steps. First, the statistical information within the primary investigation becomes converted to a common metric for comparison. The formulas for such conversions are well established and present few unique problems (Hunter & Schmidt, 1991; Rosenthal, 1984).[3] The metric chosen for this study was the correlation coefficient. The correlation represents a metric easy to transform and interpret. Statistical information with each study was corrected for artifacts of measurement such as attenuated measurement or dichotomization of independent or dependent variables (Hunter & Schmidt, 1991). A complete listing of the studies and the correlations appears in Table 1. Table 2 contains the coded features for each individual study and individual estimates for features where necessary.

Second, the separate estimates from each report must be averaged. The averaging process uses a weighted average that considers the sample size of the study. The procedures for weighted averaging are standard and used by almost all methods of meta-analysis (Bangert-Drowns, 1986; Hedges & Olkin 1985; Hunter & Schmidt, 1991; Rosenthal, 1984; Wolf, 1986).

The final step involves testing the average estimate for homogeneity. The test assumes that the average correlation comes from a single distribution of effect sizes normally distributed. The differences in effects (correlations) should only occur because of random sampling error. If that is true, then the chi-square statistic will not be significant, an insignificant chi-square suggests homogeneity. This method of meta-analysis has been called variance-centered method of meta-analysis (Bangert-Drowns, 1986).

The follow-up test for moderators based on a heterogeneous finding for an average correlation requires two complete steps to account successfully for any heterogeneity found (Hall & Rosenthal, 1991). First, there should be within group homogeneity for a successful moderator. If the variable is categorical, each category should demonstrate homogeneity within categories. This homogeneity shows that the averages for each level represent averages based on individual correlations that differ only because of sampling error. The second requirement is that there should exist significant differences in the mean correlations between groups. If the groups are different then the correlations for each group should be different. This process represents the classic homogeneity within groups and heterogeneity between groups that supports the use of analysis of variance techniques.

More sophisticated techniques involve the use of analysis of variance on the findings of à meta-analysis (Allen, 1989) or the use of multiple regression to test the contribution of moderators (Dindia & Allen, 1992). This analysis contained relatively few studies and not equally distributed across the categories, therefore no such analysis was conducted. The use of ANOVA or multiple regression in the earlier cited meta-analyses involved over 200 investigations, and even then some cells were combined or eliminated due to small cell size. The possibility of interactions between moderators remains untested.

RESULTS

Overall

The overall average correlation (ave $r = .107$, $k = 46$, $N = 64,925$) demonstrates a positive relationship between teaching evaluation scores and research productivity. The relationship observed was not based on a homogeneous set of correlations ($X^2_{(45)} = 117.94$, $p < .05$). This indicates that the average effect should be interpreted with caution since at least one moderating variable probably exists. Due to the nature of the data sets (one study with an extremely large sample size, Faia = 53,034) and several data sets included as a zero correlation (Cornwell, 1974; Grant, 1971; Lasher & Vogt, 1974; Plant & Sawrey, 1970; Ratz, 1975; Teague, 1981; Voeks, 1962) as well as those with significant correlations but no directions reported (Cornwell, 1974; Lasher & Vogt, 1974), a series of subsidiary analyses tested whether the coding of the studies can offer a sufficient explanation for the heterogeneity.[4]

Year of the Study as a Moderator

A correlation was calculated between the size of the correlation and the year of the study. A positive correlation indicates that the association between teaching evaluations and research productivity is growing. A negative correlation shows that the association between research productivity and teaching evaluations diminishes with time. The analysis shows that for both the sample size weighted (ave r = −.08) and unweighted analysis (ave r = −.04) a small negative correlation between the size of the association of the year of data collection exists. Basically, little change occurs over time. A secondary analysis examined whether the correlations changed over the various decades. The studies were separated into four groups: (a) before 1960, (b) 1960's, (c) 1970's, and (c) post-1980. The groups show that the general trend of the correlations is negative from a high correlation in the pre-1960 studies (ave r = .209, k=3, N=673, $X^2_{(2)}$= 23.98), to a smaller correlation in the 1960's decade (ave r = .095, k=13, N=3030, $X^2_{(12)}$=22.16), to a slightly larger correlation in the 1970's decade (ave r = .112, k=24, N=58,375, $X^2_{(23)}$=32.99) and the smallest average correlation occurring in the post-1980 studies (ave r = .068, k=5, N=1798, $X^2_{(4)}$=9.116). Unfortunately, not all the groups are homogeneous so this solution fails to act as a sufficient moderator.

Evaluation Method for Teaching

One issue is whether the method used to evaluate the quality of teaching acted as a moderator in the results. The five methods of evaluating the quality of instruction were: (a) student evaluations, (b) peer evaluations, (c) teaching awards, (d) amount of teaching related activities, and (e) combinations of methods. The average correlation for student evaluations (ave r = .082, k=37, N=11,177) was positive and heterogeneous ($X^2_{(36)}$= 86.65, p < .05). This indicates that there still exists a moderating condition within this group.

The peer evaluations were also heterogeneous ($X^2_{(5)}$=19.76, p < .05) and demonstrated a positive correlation (ave r = .320, k=6, N=685). Faculty peers provided a larger correlation between teaching and research productivity than did the average student evaluation correlation. Some type of halo effect may exist for faculty rating other faculty, or possibly indicate that one professional judgment of another professional stems from an understanding of the content and technique beyond that of a naive student observer.

The teaching awards create a positive correlation between winning a teaching award and research productivity (ave r = .110, k=5, N=53,337) and the effect was homogeneous ($X^2_{(4)}$=2.70, p > .05). For the teaching related activities (ave r = .325, k=1, N=174) and combinations of ratings (ave r = .199, k=1, N=27) only one study used each method and no homogeneity test can be conducted.

This moderator fails to act as a sufficient set of groupings to permit interpretation of the data. It should be noted that for all groups the correlation was positive and the absolute magnitude not that dissimilar when multiple studies appeared in the group.

Method of Measuring Research Productivity

There were eight different methods used to measure research productivity: (a) number of publications, (b) grants awarded, (c) number of citations, (d) peer rating, (e) time spent on research, (f) research awards won, (g) combination of grants and publications, and (h) research creativity. The largest number of studies (k=31) examined the number of publications. This method of measuring research productivity was positively correlated (ave r = .109, N=62,507) with teaching effectiveness but came from a heterogeneous sample of correlations ($X^2_{(30)}$=64.75, p < .05).

The number of grants awards generates a slightly higher correlation (ave r = .173, k=2, N=211) with a homogeneous sample ($X^2_{(1)}$=0.60, p > .05). Given the small number of studies and total sample size, any definitive conclusions remain difficult to draw from this data. Homogeneity also existed for the relationship between the number of citations and teaching evaluations (ave r = -.032, k=5, N=1036) ($X^2_{(4)}$=6.85, p > .05). This constitutes the only average correlation that was negative but the small number of studies and sample size make it difficult to draw firm conclusions.

The peer rating of research demonstrates a typical relationship with teaching evaluations (ave r = .124, k=7, N=858) based on a homogeneous sample ($X^2_{(6)}$=6.40, p > .05). The sample sizes and correlations for the research time group (ave r = .000, k=2, N=1558, zero variance) and the combination grants and publications group (ave r = .030, k=2, N=160, $X^2_{(1)}$=0.32, p > .05) were small. The research awards group (ave r = .000, k=1, N=15) and research creativity group (ave r = .540, k=1, N=86) contained only one study apiece.

Methods of measuring research productivity as a moderator fail to account for the available heterogeneity. The number of studies available remained too small in many categories to justify their existence and the only category with a large number of studies exhibits significant heterogeneity.

CONCLUSIONS

The results show a positive correlation between teaching effectiveness and research productivity (ave r = .107). The correlation indicates that as either teaching effectiveness or research productivity increases the other variable does as well. Only a correlation exists, no ability to evaluate any causality between the features is possible given the restricted set of conditions of the data. Such an interpretation requires the introduction of some type of theoretical mechanism that translates the increasing value of one variable into an effect that increases the value of the other variable. While arguments exist about the nature of the connection, the exact causal connection remains unclear. The ability to infer that a causal connection exists between research productivity and teaching effectiveness does not exist within this report. While such a connection is a necessary condition for an interpretation of causality, the correlation is not a sufficient condition.

There are two considerations that need to be addressed when considering the impact of the connection observed. First, the average correlation comes from a sample of heterogeneous effects, therefore any interpretation must be cautious. The inability to generate a homogeneous solution using the moderators provides some uncertainty about the ability to generalize the average observed. The only exception to the rule that all average correlations observed were positive or zero occurred for the number of citations used for a measure of research productivity. Most likely the "true" moderator variable would separate one smaller positive effect from a larger positive effect. While this separation is important it is a distinction between magnitudes of positive correlations not the direction of the correlation. It seems warranted to expect, for example, that the sheer number of publications is positively correlated with teaching evaluations, even when considering any moderating variables.

The second issue considers the importance of the size of the average correlation. Many would consider a correlation of .10 as small and unimportant, contributing to little of the variance. Abelson (1985) points out that there exists much misconception about the importance of what amounts to small effects. For example, the difference between a .200 and a .300 hitter in baseball for any one at bat is .00317 (using omega squared) and the effect observed in this meta-analysis is far larger than that difference.

The importance of the connection can be illustrated in terms of utilizing a Binomial Effect Size Display (BESD). Suppose we have 200 faculty members and 100 do research and 100 do not. The overall teaching evaluation average across all 200 faculty is 50 with a standard deviation of 10. We have their teaching evaluations and we find the correlation between conducting research and teaching evaluations is .10 (the same as the average in this meta-analysis). The mean for the research faculty for teaching evaluations would be 51 and for the nonresearch faculty the mean would be 49 (this is based on d= .20, which is the same as the correlation of .10, d = (difference between means or 2) / (standard deviation or 10)). This is not a large difference but suppose we want the top 16% of the teaching (one standard deviation above 50 or a cutoff score of 60) you would find that 18.41% of the research faculty above that score while only 13.57% of the nonresearch faculty above that score. Suppose we want the very best teachers and use a three positive standard deviation cutoff score (score of 80) we would find that 0.19% of the research faculty and 0.10% of the nonresearch faculty achieving that goal. For the last severe cutoff score there is an almost 2 to 1 advantage in the probability of a research faculty member attaining a high ranking when compared to a nonresearch faculty member. Table 3 provides an extended statistical summary of the various features of the impact of a correlation of .10 on the comparisons.

The practical implications of the finding deserve some consideration. While research is not a perfect indication of high quality teaching, clearly productive research is not inconsistent with quality teaching. More than likely there is at some point a level of diminishing returns where research efforts operate to reduce the quality of teaching but that point is not developed in this data. The data do clearly support the idea that research productivity and quality teaching are not contradictory goals, the degree to which they are compatible or complementary goals could still be argued.

One interpretation implies the existence of some third factor underlying both positive teaching evaluations and research productivity. Some underlying personality or professional characteristic creates an outcome that favors both quality teaching and high research productivity. One interesting issue developed in an early study (Maslow & Zimmerman, 1956) was the high correlation between research creativity and teaching evaluation. If one were to list those features contributing to successful research (creativity, hard work, etc.) and compare them to those features contributing to successful teaching (hard work, ability to explain) we might find many features in common.

A second explanation might be that the nature of the teaching evaluations interacts with time spent on teaching. Suppose that researcher's classes become routine standardized exercises that vary little from year to year and the faculty member gives high grades. The researcher might well be evaluated positively, while simultaneously providing little inspiration to improve the students. The relationship between positive overall evaluations and effectiveness as measured by actual learning and understanding material is unclear. If the argument that research degrades quality teaching the argument turns on some feature of the methodological aspects of evaluation being associated with research productivity.

Future research should consider the issues about the impact of research and the actual process of learning. Communication scientists could make an impact by examining how researchers interact within the classroom and whether that interaction is different from nonpublishing faculty. Teaching is an interactive process between student and teacher, where the teacher provides knowledge as well as instills curiosity within the students. One issue of Communication Education (edited by Lawrence Rosenfield in 1994) contains stories about the nature of how teachers interacted with students to create a positive influence. None of the stories or commentaries provides a connection between good research and good teaching.

The ongoing debate over what duties a faculty member should be responsible for continues to receive attention. This summary provides some information about one relationship, between positive teaching evaluations and research productivity, by finding a positive relationship. The results provide a preliminary conclusion that there exists a small positive relationship between teaching and research. The findings do not end the argument about the nature of the trade-off but they do provide evidence that the trade-off is not inevitable. Our departments need to consider this aspect of the relationship before establishing standards for hiring, merit, promotion, and tenure.

FOOTNOTES

[1]The arguments surrounding the measurement and documentation of effective as well as positively evaluated teaching deserve consideration but are beyond the scope of this report. For discussion of these issues the reader is referred to other relevant meta-analyses (Feldman, 1976a; 1976b; 1977; 1978; 1979; 1983; 1984; 1986).

[2]Several data sets were available in multiple manuscripts (Aleamonia and Yimer, 1972; 1973; 1974; Freedman, Stumpf & Aguanno, 1976; Friedrich & Michalak, 1983; Hoffman 1984a; 1984b; Michalak & Friedrich, 1981; Stumpf, Freedman, & Aguanno, 1979; Wood, 1977, 1978). Several manuscripts only offered reviews of the available literature (Blackburn, 1972; Braxton & Bayer, 1986; Schachter, 1991) useful for bibliographic purposes but did not contribute data sets. Data reported by Aiken (1975) were not included because no sample size was reported.

[3]There were several discrepancies between the correlations Feldman (1987) reports using and what are used in this report. For example he provides no estimate for Ahern (1969) or Goldsmid, Gruber, and Wilson (1978) although a significant correlation exists, this report estimates a correlation that would be minimally significant given the available sample size. Feldman reports a correlation of .30 which is the value of the student evaluations, the correlation used here is .255 which is the average of the student and the peer evaluations (.210), this procedure also accounts for the discrepancy between the Friedrich and Michalak (1983) correlation. Feldman also reports a sample size of 211 for the Harry and Goldner (1972) study, the indication on table 2 of the article is that every correlation has a different relevant sample size, the relevant sample size for this correlation is not 211 but 77. The Hayes (1971) study had multiple sample sizes for each method of teaching

evaluations, this report uses the average sample size. Feldman did not consider the Hoyt (1974) or Rossman (1976) measures as an indication of teaching evaluation since it asked students to rate how much they had improved because of the course in the subject matter. For this summary the measure was considered appropriate as a student evaluation of a course. Several estimates are larger in this report than those in the Feldman report (see for example Marquardt, McGann, and Jokubauskas study (.280) verus that provided by Feldman (.250)) because of application of the correction for measurement error due to attenuation. Comparing the uncorrected estimates of this report to the uncorrected estimates provided by Feldman finds complete correspondence except where noted. Feldman provides no estimate for the Riley, Ryan, and Lifshitz (1950) study because no calculation is possible. This investigation calculates percentages of published versus unpublished faculty obtaining superior teaching ratings. This permits the calculation of a d statistic using the table in Glass, McGaw, and Smith (1981). In estimating the effect for Stavridis, Feldman used only a few of the student evaluation items, this report used a much later percentage of the items and the correlation changed from .240 to .163 (it should be noted that had the same items been utilized the correlations would have been identical).

[4]The subsidiary analyses demonstrate that despite the procedure used, the same basic results were obtained. For example, the analysis conducted by excluding the Faia study reveals the same basic results (ave r = .094, k = 46, N = 11,933 with homogeneity results of ($X^2_{(45)}$ = 115.438, p < .05). Conducting the analysis without studies whose values could not be calculated exactly and where estimates of zero were entered (Cornwell, Teague, Voeks, Grant, Lasher, Ratz, Plant) demonstrates no divergent results (ave r = .111, k = 39, N = 63,013 with homogeneity test $X^2_{(38)}$

= 87.847). Another analysis conducted removing the studies with valued entered at zero and entering either minimal positive values for studies with significant but nondirectional findings (Lasher and Cornwell) utilizing the procedures outlined by Boster and Mongeau (1984) reveals no difference in results (ave r = .112) or when using minimal negative values either (ave r = .110). Except where reported in the text, all subsequent moderator analyses conducted the same procedure for inclusion and considering the possible influences of coding procedures. Except where indicated in the text no differences existed based on this analysis. A complete detailed series of the analyses is available from the author.

REFERENCES

Abelson, R. (1985). A variance explanation paradox: When a little is a lot. Psychological Bulletin, 97, 129-133.

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 1963-1964. Unpublished doctoral dissertation, University of Massachusetts, Amherst Massachusetts.

Aiken, L. (1975). Procedures and problems in designing a college course evaluation questionnaire. College and University, 50, 247-253.

Aleamonia, L., & Yimer, M. (1972). An investigation of the relationship between colleague rating, student rating, research productivity, and academic rank in rating instructional effectiveness. Office of Instructional Resources, Report #338, University of Illinois, Urbana-Champaign, Illinois.

Aleamonia, L., & 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, 274-277.

Aleamonia, L., & Yimer, M. (1974). Graduating senior ratings relationship to colleague rating, student rating, research productivity and academic rank in rating instructional effectiveness. Research Report #352, Office of Instructional Resources, Measurement and Research Division, University of Illinois, Urbana-Champaign, Illinois.

Allen, M. (1989). A comparison of self report, observer, and physiological assessments of public speaking anxiety reduction techniques using meta-analysis. Communication Studies, 40, 127-139.

Allen, M. (1993). Determining the persuasiveness of message sidedness: A prudent note about utilizing research summaries. Western Journal of Communication, 57, 98-103.

Allen, M. (1991). Meta-analysis comparing the effectiveness of one and two-sided messages. Western Journal of Speech Communication, 55, 390-404.

Allen, M., Hunter, J., & Donohue, W. (1989). Meta-analysis of self-report data on the effectiveness of public speaking anxiety treatment techniques. Communication Education, 38, 54-76.

Allen, M., & Preiss, R. (1990). Using meta-analyses to evaluate curriculum: An examination of selected college textbooks. Communication Education, 38, 103-116.

Allen, M., & Preiss, R. (1993). Replication and meta-analysis: A necessary connection. Journal of Social Behavior and Personality, 8, 9-20.

Bangert-Drowns, R. (1986). Review of developments in meta-analytic method. Psychological Bulletin, 99, 388-399.

Barker, L, Hall, R., Roach, D., & Underberg, L. (1979). An investigation of the quantity of articles produced in the communication discipline: 1970 through 1978. Association for Communication Administration Bulletin, 30, 18-22.

Barker, L., Hall, R., Roach, D., & Underberg. L. (1980). An investigation of articles produced in the communication discipline by institution: A journal-by-journal, year-by-year analysis. Association for Communication Administration Bulletin, 34, 37-48.

Barker, L., Hall, R., Roach, D., & Underberg, L. (1981). An investigation of articles produced in the communication discipline: An institutional analysis on the basis of highest degree conferred, part III. Association for Communication Administration Bulletin, 37, 34-39.

Barker, L., Ray, V., Watson, K., & Hall, R. (1988). An investigation of the quantity of articles produced in the communication discipline: An institutional analysis on the basis of the highest degree conferred. Association for Communication Administration Bulletin, 65, 31-38.

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

Beaman, A. B., Cole, C. M., Preston, M., Klentz, B., and Steblay, N. M. (1988). Fifteen years of foot-in-the-door research: A meta-analysis. Personality and Social Psychology Bulletin, 9, 181-196.

Blackburn, R. (1972). Tenure: Aspects of job security on the changing campus. Research Monograph 19. Southern Regional Education Board, Atlanta, Georgia.

Bloom, A. (1987). The closing of the American mind. New York: Simon and Schuster.

Boster, F.J., & Mongeau, P. (1984). Fear-arousing persuasive messages. In R. Bostrom (Ed.), Communication Yearbook 8 (pp. 330-375). Newbury Park, CA: Sage.

Braunstein, D., & Benston, G. (1973). An assessment of faculty performance: Some correlates between administrator, colleague, student, and self-ratings. Sociology of Education, 48, 242-256.

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

Braxton, J., & Bayer, A. (1986). Assessing faculty scholarly performance. Measuring Faculty Research Performance: New Direction for Institutional Research, 50, 25-42.

Bresler, J. (1968). Teaching effectiveness and government awards. Science, 160, 164-167.

Blair, C., Brown, J., & Baxter, L. (1994). Disciplining the feminine. Quarterly Journal of Speech, 80, 383-409.

Burgoon, M. (1988). Extramural funding or extracurricular research: That is the choice. A research editorial. Western Journal of Speech Communication, 52, 252-258.

Burgoon, M. (1989). Instruction about communication: On divorcing Dame Speech. Communication Education, 38, 303-308.

Centra, J. (1983). Research productivity and teaching effectiveness. Research in Higher Education, 18, 379-389.

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

Cornwell, C. (1974). Statistical treatment of data from student teaching evaluation questionnaires. Journal of Chemical Education, 51, 155-160.

Dent, P., & Lewis, D. (1976). The relationship between teaching effectiveness and measures of research quality. Educational Research Quarterly, 1, 3-16.

Dillard, J., & Hale, J. (1993). Prosocialness and sequential request compliance techniques: Limits to the foot-in-the-door and the door-in-the-face? Communication Studies, 43, 220-232.

Dillard, J. P., Hunter, J. E., & Burgoon, M. (1984). Sequential-request persuasive strategies: Meta-analysis of foot-in-the-door and door-in-the-face. Human Communication Research, 10, 461-488.

Dindia, K., & Allen, M. (1992). Sex differences in self-disclosure: A meta-analysis. Psychological Bulletin, 112, 106-124.

Erickson, K., Fleuriet, C., & Hosman, L. (1993). Prolific publishing: Professional and administrative concerns. Southern Communication Journal, 58, 328-338.

Faia, M. (1976). Teaching and research: Rapport or misalliance. Research in Higher Education, 4, 235-246.

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

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

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

Feldman, K. (1978). The significance of circumstances for college students' ratings of their teachers: What we know and what we don't. Research in Higher Education, 9, 199-242.

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

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

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

Feldman, K. (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, 139-213.

Feldman, K. (1987). Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness: A review and exploration. Research in Higher Education, 26, 227-298.

Fern, E. F., Monroe, K. B., & Avila, R. A. (1986). Effectiveness of multiple request strategies: A synthesis of research results. Journal of Marketing Research, 23, 144-152.

Freedman, R., Stumpf, S., & Aguanno, J. (1979). Validity of the course-faculty instrument (CFI): Intrinsic and extrinsic variables. Educational and Psychological Measurement, 39, 153=159.

Frey, P. (1978). A two-dimensional analysis of student ratings of instruction. Research in Higher Education, 54, 145-163.

Friedrich, R., & Michalak, S. (1983). Why doesn't research improve teaching? Some answers from a small liberal arts college. Journal of Higher Education, 54, 145-163.

Glass, G., McGaw, B., & Smith, M. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.

Goldsmid, C., Gruber., J., & Wilson, E. (1977). Perceived attributes of superior teachers (PAST): An inquiry into the giving of teaching awards. American Educational Research Journal, 14, 423-440.

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

Hall, J., & Rosenthal, R. (1991). Testing for moderator variables in meta-analysis: Issues and methods. Communication Monographs, 58, 437-448.

Hammond, P., Meyer, J., & Miller, D. (1969). Teaching versus research: Sources of misperceptions. Journal of Higher Education, 40, 683-690.

Harry, J., & Goldner, N. (1972). The null relationship between teaching and research. Sociology of Education, 45, 47-60.

Hayes, J. (1971). Research, teaching, and faculty fate. Science, 172, 227-230.

Hedges, L., & Olkin, I. (1985). Statistical methods for meta-analysis. Orlando, FL: Academic Press.

Hicks, R. (1974). The relationship between publishing and teaching effectiveness. California Journal of Educational Research, 25, 140-146.

Hickson, M. (1990). Profiling the chairs of prolific speech communication departments. Association for Communication Administration Bulletin, 73, 4-14.

Hickson, M., Stacks, D., & Amsbary, J. (1989). An analysis of prolific scholarship in speech communication, 1915-1985: Toward a yardstick for measuring research productivity. Communication Education, 38, 230-236.

Hickson, M., Stacks, D., & Amsbary, J. (1992). Active female scholars in communication: An analysis of research productivity, II. Communication Quarterly, 40, 350-356.

Hickson, M., Stacks, D., & Amsbary, J. (1993). Active prolific scholars in communication studies: Analysis of research productivity, II. Communication Education, 42, 224-233.

Hoffman, R. (1984a). Correlates of faculty performance. College Student Journal, 18, 164-168.

Hoffman, R. (1984b). An assessment of the teaching, research, and service function of a college faculty. Journal of Research and Development in Education, 17, 51-54.

Hoyt, D. (1974). Interrelationships among instructional effectiveness, publication record, and monetary reward. Research in Higher Education, 2, 81-88.

Hoyt, D., & Spangler, R. (1976). Faculty research involvement and instructional outcomes. Research in Higher Education, 4, 112-122.

Hunter, J., Hamilton, M., & Allen, M. (1989). The design and analysis of language experiments in communication. Communication Monographs, 56, 341-363.

Hunter, J., & Schmidt, F. (1991). Methods of Meta-Analysis. Beverly Hills, CA: Sage.

Jauch, L. (1976). Relationships of research and teaching: Implications for faculty evaluation. Research in Higher Education, 4(2), 112-122.

Kim, M., & Hunter, J. (1993a). Attitude-behavior relations: A meta-analysis of attitudinal relevance and topic. Journal of Communication, 43, 101-142.

Kim, M., & Hunter, J. (1993b). Relationships among attitudes, behavioral intentions, and behavior: A meta-analysis of past research, Part 2. Communication Research, 20, 331-364.

Kurland, N. (1961). Teaching and publication. Journal of Higher Education, 32, 169-171.

Lasher, H., & Vogt, K. (1974). Student evaluations: Myths and realities. Improving College and University Teaching, 22, 267-269.

Linsky, A., & Straus, M. (1975). Student evaluations, research productivity, and eminence of college faculty. Journal of Higher Education, 46, 89-102.

Louis, J. (1992). The compatability of teaching and research. Journal of Chemical Education, 69, 603.

Marquardt, R., McGann, A., & Jakubauskas, E. (1975). Academic clients, scholarly contributions, faculty compensation, and performance criteria, or "Shouldn't we take one more look before we leap...?" AACSB Bulletin, 12, 13-17.

Martin, T., & Berry, K. (1961). The teaching-research dilemma: Its sources in the university setting. Journal of Higher Education, 32, 691-703.

Maslow, Z., & Zimmerman, W. (1956). College teaching ability, scholarly activity, and personality. Journal of Educational Psychology, 47, 185-189.

McCullagh, R., & Roy, M. (1975). The contribution of noninstructional activities of college classroom teacher effectiveness. Journal of Experimental Education, 44, 61-70.

McDaniel, E., & Fedlhusen, J. (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.

Michalak, S., & Friedrich, R. (1981). Research productivity and teaching effectiveness at a small liberal arts college. Journal of Higher Education, 52, 578-597.

Mongeau, P. (1994). Another Look at Fear-Arousing Persuasive Appeals. In M. Allen and R. Preiss (Eds.), Persuasion: Advances Through Meta-Analysis (pp. 75-100). Dubuque, IA: Brown and Benchmark.

O'Keefe, D. (1993). The persuasive effects of message sidedness variations: A cautionary note concerning Allen's (1991) meta-analysis. Western Journal of Communication, 57, 87-97.

Osborn, M., & Osborn, S. (1994). Public speaking (3rd ed.). Boston, MA: Houghton Mifflin Co.

Plant, W., & Sawrey, J. (1970). Student ratings of psychology professors as teachers and research involvement of the professors rated. The Clinical Psychologist, 23, 15-16, 19.

Preiss, R., & Allen, M. (1994). Prospects and precautions in the use of meta-analysis. In M. Allen and R. Preiss (Eds.), Persuasion: Advances through meta-analysis (pp. 1-33). Dubuque, IA: Brown and Benchmark Publishers.

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

Richardson, P., Parker, R., & Udell, G. (1992). Does research enhance or inhibit teaching? An exploratory study. Journal of Education for Business, 92, 79-83.

Riley, J., Ryan, B., & Lifshitz, M. (1950). The student looks at his teacher: An inquiry into the implications of student ratings at the college level. New Brunswick, NJ: Rutgers University Press.

Rodden, J. (1993). Field of dreams. Western Journal of Communication, 57, 111-138.

Root, L. (1987). Faculty evaluation: Reliability of peer assessments of research, teaching, and service. Research in Higher Education, 26, 71-84.

Rosenfeld, L. (Ed.) (1993). Special Issue: When teaching "works": Stories of communication in education. Communication Education, 42, 277-401.

Rosenthal, R. (1984). Meta-analytic procedures for social research. Beverly Hills, CA: Sage.

Ross, R. (1992). Speech communication: The speechmaking process (9th ed.). Englewood Cliffs, NJ: Prentice Hall.

Rossman, J. (1976). Teaching, publication, and rewards at a liberal arts college. Improving College and University Teaching, 24, 238-240.

Rushton, L., Murray, H., & Paunonen, S. (1983). Personality, research creativity, and teaching effectiveness in university professors. Scientometrics, 5, 93-116.

Schachter, H. (1991). Teaching versus research: A 1910 perspective. College Teaching, 39, 85-86.

Sheppard, B., Hartwick, J., & Warshaw, P. (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. Journal of Consumer Research, 5, 325-343.

Siegfried, J., & White, K. (1973). Teaching and publishing as determinants of academic salaries. Journal of Economic Education, 4, 28-42.

Smith, P. (1961). Teaching, research, and publication. Journal of Higher Education, 32, 199-205.

Stacks, D., & Hickson, M. (1983). An analysis of doctoral degree-granting institutions and number of articles published by their graduates. Association for Communication Administrative Bulletin, 43, 47-52.

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

Stavridis, P. (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., Freedman, R., & Aguanno, J. (1979). A path analysis of factors often found to be related to student ratings of teaching effectiveness. Research in Higher Education, 11, 111-123.

Sutton, S.R. (1982). Fear-arousing communications: A critical examination of theory and research. In J.R. Eisne (Ed.), Social psychology and behavioral medicine. New York: John Wiley.

Sykes, C. Profscam: Professors and the demise of higher education. Washington, D.C.: Regnery Gateway.

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

Usher, R. (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, Gainesville, Florida.

Voeks, V. (1962). Publications and teaching effectiveness. Journal of Higher Education, 33, 212-218.

Watson, K., Barker, L., Ray, V., & Hall, R. (1988). An investigation of quantity of articles produced in the communication discipline by institution: 1980-1985. Association for Communication Administration Bulletin, 63, 85-90.

Winer, B. (1971). Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill.

Wolf, F. (1986). Meta-analysis: Quantitative methods for research synthesis. Newbury Park, CA: Sage.

Wood, P. (1977). The description and evaluation of a college department's faculty rating system. Paper presented at the annual meeting of the American Educational Research Association, New York.

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

Educational Research Association, Toronto, Canada.

Wood, P., & DeLorne, C. (1976). An investigation of the relationship among teching evaluation, research, and ability. Journal of Economic Education, 7, 77-80.

Woodburne, L. (1952). The qualifications of superior faculty members. Journal of Higher Education, 23, 377-382.

Table 1

Effect Sizes Relating Research Productivity and Teaching Effectiveness

Author[1]	Date[2]	r[3]	N
Ahern	1969	.238	75
Aleamoni (1)	1973	.000	360
(2)		.033	28
Bausell	1972	.061	105
Braunstein	1973	.040	349
Braxton	1983	.325	174
Bresler	1968	.227	106
Centra (1)	1983	.099	2,968
(2)		.071	1,623
Clark	1973	.255	45
Cornwell	1974	.000	70
Dent	1976	.022	90
Faia	1976	.110	53,034
Freedman	1979	.242	129
Frey	1978	.070	42
Goldsmid	1977	.172	90
Grant	1971	.000	685
Harry	1972	.190	77
Hayes	1971	.210	250
Hicks	1974	.192	459
Hoffman	1984	-.250	65
Hoyt	1974	.086	173
Hoyt	1976	.170	183
Lasher	1974	.000	873
Linsky	1975	.009	1,091
Marquardt	1975	.286	91
Maslow	1956	.640	86
McCullagh	1975	.045	52
McDaniel	1970	.043	76
Michalak	1981	.260	86
Plant	1970	.000	32
Ratz	1975	.000	15
Richardson	1992	.260	67
Riley	1950	.220	389
Root	1987	.199	27
Rossman	1976	.327	122
Rushton	1983	-.066	52
Siegfried	1973	.039	45
Stallings (1)	1970	.260	128
(2)		.105	121
Stavridis	1972	.163	32
Teague	1981	.000	16
Usher	1966	.230	26

Table 1 (Cont.)

Voeks	1962	.000	198
Wood	1976	.395	69
Wood	1978	.023	22

[1]First author listed, see References for complete citation

[2]Date listed is date of publication not data set

[3]Correlation reported is corrected correlation averaged across multiple measures. When the correlation differs from that reported by Feldman (1987) an explanation is provided in Footnote 2.

Table 2

Methods of Measuring Research Productivity and Teaching Effectiveness

		Method of Measuring	
Author[1]	Date[2]	Research[3] Productivity	Teaching[3] Effectiveness
Ahern	1963-64	# of pub	awards
Aleamoni (1)	1969-70	# of pub	student
(2)	1969-1970	# of pub	peer
Bausell	1969-70	# of pub (.041) grants (.119)	student
Braunstein	1968-69	peer	student
Braxton	1979	# of pub	activities
Bresler	1965	grants	student
Centra (1)	1979	# of pub	student
(2)	1980	# of pub	student
Clark	1968	# of pub	peer (.210) student (.300)
Cornwell	1972	peer	student
Dent	1973	# of pub (.051) citations (-.008)	student
Faia	1973	# of pub	awards
Freedman	1979	# of pub	student
Frey	1975-76	citations	student
Goldsmid	1972-74	# of pub	awards
Grant	1968	time	student
Harry	1968-69	# of pub	student
Hayes	1967-69	# of pub	peer (.289, 318) student (.073, 183)
Hicks	(1974)	# of pub	student
Hoffman	1980	# of pub	student
Hoyt	1969	# of pub	student
Hoyt	1972-73	peer	student
Lasher	1969	time	student
Linsky	(1955-75)	# of pub (.040, 1422) citations (-.050, 766)	student
Marquardt	1972	peer	student
Maslow	1943-46	creativity	student (.510) peer (.77)
McCullugh	1971-72	# of pub	student
McDaniel	(1970)	# of pub	student
Michalak	1977-78	# of pub (.320) citations (.200)	peer
Plant	(1970)	pub and grant	student
Ratz	1969-75	awards	student
Richardson	(1992)	# of pub	student

Table 2 (Cont.)

Riley	1947	# of pub	student
Root	1985	peer	combination
Rossman	1969	peer	awards (.190) peer (.230)
Rushton	1974-79	# of pub (.149) citations (-.280)	student
Siegfried	1970-71	# of pub	student
Stallings (1)	1965-66	# of pub	student
(2)	1967	# of pub	student
Stavridis	1972	# of pub	student
Teague	(1981)	# of pub	awards
Usher	1965	peer	student
Voeks	1948-52	# of pub	student
Wood	1971-73	# of pub	student
Wood	1974-77	pub and grant	student

[1]First author listed, see References for complete citation

[2]Date listed is date of actual data collection not publication date, if date is in parantheses that indicates publication date because actual data collection date not available.

[3]First number in parathesis is correlation, second number is sample size if different from overall. No entry means overall correlation is based on this value.

Table 3

Binomial Effect Size Display for Interpreting Results

The following assumes that r = .10 and a scale with a mean = 50, standard deviation = 10 and that each faculty group (research and nonresearch) is equal in number.

Cutoff score	Percentage of faculty past the cutoff score		Ratio of research to nonresearch faculty	Percentage chance faculty member past cutoff score is research faculty member
	research	nonresearch		
Greater than the mean Above average	55.00%	45.00%	1.22 to 1	55%
Greater than one standard deviation "Excellence"	18.41%	13.57%	1.36 to 1	58%
Greater than two standard deviations "Outstanding"	2.87%	1.79%	1.60 to 1	62%
Greater than three standard deviations "Teacher of the year"	0.19%	0.10%	1.90 to 1	66%