DOCUMENT RESUME

ED 421 470 SP 038 039

AUTHOR Burns, Candace W.
TITLE Peer Evaluation of Teaching: Claims vs. Research.
PUB DATE 1998-00-00
NOTE 24p.
PUB TYPE Information Analyses (070)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Faculty; Elementary Secondary Education; Evaluation Methods; Higher Education; *Peer Evaluation; *Student Evaluation of Teacher Performance; Teacher Competencies; *Teacher Evaluation; Teachers; Teaching Skills

ABSTRACT

This article examines evidence supporting the validity of peer evaluations of teaching, focusing on relationships between peer evaluations and student evaluations of teaching. The article discusses conditions in which peer and student ratings are comparable, beginning with studies that K. A. Feldman included in his 1989 meta-analysis and concluding with studies post-dating Feldman's analysis, which include expanded evaluations of teaching. In particular, the article notes that as the specificity of information that peers receive about faculty members' teaching increases, agreement with student ratings decline. Conversely, peer ratings based on general, impressionistic information tend to agree with student ratings. Though there is some evidence of the potential for peers to aid in teacher evaluation, the evidence for peers' effectiveness in broadened evaluative roles is scant and inconsistent. Given mixed and sometimes weak evidence for the validity of peer ratings, the article calls for caution regarding roles that peers should assume in evaluating teaching. (Contains 25 references.) (SM)

Running head: PEER EVALUATION

Peer Evaluation of Teaching: Claims vs. Research

Candace W. Burns

University of Arkansas-Little Rock

Abstract

This article examines evidence for the validity of peer evaluations of teaching. Because the bulk of validity studies have compared peer and student evaluations; and because the validity of student evaluations has received intensive scrutiny and considerable empirical backing (Marsh, 1987; Scriven, 1995), this article focuses on conditions that appear to promote peer-student agreement. In particular, it notes that as the specificity of information that peers receive about faculty members' teaching increases, agreement with student ratings declines. Conversely, peer ratings based on general, "impressionistic" information, tend to agree with student ratings. Given mixed and sometimes weak evidence for the validity of peer ratings, this article calls for caution regarding roles that peers should assume in evaluating teaching.

Peer Evaluation of Teaching: Claims vs. Research

Marsh (1987) noted that "while extensive lists of alternative indicators of effective teaching are proposed, few are supported by systematic research, and none are as clearly supported as current students' evaluations of teaching" (p. 369). Nevertheless, such lists suggest that colleagues might be uniquely suited to judge a variety of aspects of their peers' teaching, including (a) course-related information (organization, selection and mastery of content, and appropriateness of objectives); (b) instruction-related information (appropriateness of materials, media use, exams, methodologies, etc.); and (c) student-centered information (commitment to student learning and achievement and availability to students) (Cohen & McKeachie, 1980, p. 148; Cashin, 1989, pp. 3-4).

Most recently, a Carnegie survey of 865 four-year colleges (as cited in Magner,1997, p. A18) reported that while 98% used student evaluations, 62% used and 29% were considering using peer review of instruction-related information. Further, 58% used and 33% were considering using peer review of classroom teaching; and 24% used and 41% were considering using student-centered information. Despite a certain face validity for broadening the scope of peer review of teaching, the bulk of empirical research has focused on comparisons of peer and student ratings of teaching (Feldman, 1989).

Because student ratings are both widely-used and have received much empirical support, and peer review of teaching is

increasingly popular, the present article will focus on relationships between these two types of ratings. It will examine conditions in which peer and student ratings are comparable, beginning with studies that Feldman (1989) included in his meta-analysis and concluding with studies that post-date Feldman's analysis--a few of which include expanded evaluations of teaching. It will conclude with issues that faculty and administrators may wish to consider as they pursue recommendations and practices for the review of teaching.

Research Which Compares Student and Peer Ratings of Teaching

Measures of the same construct, such as effective teaching, should be related, thereby showing evidence of convergent validity. In the present case, effective teaching should be recognizable by both peers and students.

Feldman's Metaanalysis (1989)

Feldman (1989) reported the mean correlation between peer and student ratings across 14 studies as .55; however, in the individual studies that he reviewed, actual correlations ranged from .19 to .84. McMillan (1996, p. 118) described correlations between .10 to .30 as low, those between .40 and .60 as moderate, and those of .70 or greater as high. The discussion of correlations will use McMillan's descriptors.

Studies with High Current Student-Peer Correlations

Ballard, Reardon, and Nelson (1976), r=.84; McCarbery (1970), r=.84; Murray (n.d); Murray (1972), r=.82; Choy (1969), r=.78; Maslow and Zimmerman (1956), r=.69. All of these studies allowed

comparisons of peer and student ratings of overall teaching effectiveness. Peer raters received no specific information to examine, (i.e., student ratings or teaching portfolios). They were chosen as raters based on their familiarity with the faculty.

These studies used varied scales: poor vs. excellent (Ballard et al.,1976); good, average, poor (Choy, 1969); others combined questionaire items to produce a total rating (McCarbery, 1970). Some used ratings of overall teaching effectiveness, with no scale specified (Murray, n.d., & Murray, 1972).

Marsh contended that in studies where peer-student correlations were high, and in which peers did not observe classroom performance, that peers may have based their ratings on information from students (1987, p.294). One cannot rule out this explanation because these studies did not control for peer familiarity with student ratings.

In addition, rating scales in at least two of the studies (Ballard et al., 1976; Choy, 1969) may have artificially induced agreement. In the Ballard et al. study, subjects simply had to choose between ratings of poor or excellent. In the Choy study, raw data revealed that only 4 of 32 raters (student and peers) chose the category, poor.

In sum, ratings based on peer impressions of faculty teaching tend to agree with student ratings. Sometimes such agreement may have been artificially high because studies used rating scales with limited choices; moreover, they did not control for student influences on peer ratings. One may ask, however, what might

happen to peer-student agreement if peers review more information before rating teaching?

Studies with Low Current Student-Peer Correlations

Howard, Conway, and Maxwell (1985), r=.19; Centra (1975); r=.23; Aleamoni and Yimer (1973), r=.27; Wood (1978), r=.36. Unlike the peer raters (discussed in the previous section),whose ratings correlated highly with student ratings, and who received no specific information on which to base their ratings, raters in two studies in this group (Howard et al. and Centra) actually observed teaching. Similar to studies in the previous section, the Aleamoni and Yimer and Wood studies did not provide peer raters with specific information on teaching.

Howard et al. (1985) obtained ratings from current and former students, external observers, self, and peers. Briefly, their information-gathering procedures were: (a) obtaining current student ratings two weeks before the end of the semester; (b) randomly sampling former students that each instructor had taught for the past two years; (c) obtaining peer ratings based on two 50-75 minute observations by the same colleague; and (d) obtaining trained external observer ratings based on two 50-75 minute observations. All raters used the same evaluation instrument which Howard et al. reduced to a single score. All used a 9-point scale.

Howard et al. (1985) examined the impact of validation methods on the size of correlations. They attained a low correlation for peer-student evaluations (r=.19) when they used peer evaluations as sole validation criterion for student

evaluations. When they compared current student evaluations to other observations, student evaluations emerged with the greatest relationship to all and, therefore, to the construct, teaching. As example, after obtaining all ratings, Howard et al. calculated the correlation between each rating method. Then, they averaged the correlations between a particular method and each of the other methods to produce a single coefficient for that method. The average validity coefficients were: current and former students, .38 and .37 (moderate); self, .20 (low); trained observers, .12 (low); and peers, .10 (low). Howard et al. asserted that student ratings may have shown low correlations in other studies because they were validated against a criterion (peer ratings) of low validity (p. 195).

Two other demonstrations made essentially the same case. In one, former and current student ratings correlated .63 and .48 (moderate) with combined ratings from all sources, yet peer evaluations correlated only .01 (low) with combined ratings. When using confirmatory factor analysis, correlations for former and current student ratings were high (.88 and .78); and correlations for peer ratings were low (.25).

Given that peer raters in their study had more information than the raters in studies finding high student-peer correlations, it is troubing that more information did not afford better judgments. One cannot discount, however, that the observation-based peer evaluations that Howard et al. (1985) used were a weak type of peer evaluation--only two per untrained colleague

observer. Howard et al. suggested that compared to student raters, whose observations are "averaged over multiple judges," and who have over "20 times the exposure to their instructor's teaching," peer raters' accuracies probably contain more error due to "sampling bias" (p. 195). They noted that eight trained external observers did not perform very well either.

In an attempt to control for student influences on peer ratings, Centra (1975) investigated the relationship between peer-student teaching ratings in a newly-opened college. Three peers observed the instructors twice for an unspecified length of time. Both students and peer observers rated teaching using 16 items from the 39-item Student Instructional Report (SIR). Although three-quarters of the faculty (N=78) agreed to be rated by peers and students, Centra received complete data for only 54 faculty.

When calculating the correlation between peer and student ratings, Centra (1975) used all 78 faculty. The increased number of faculty included in the correlations may have inflated the correlations. Regardless, the overall correlation between peer-student teaching ratings was low--r=.23. Peer-student correlations for six specific teaching dimensions were moderate (but possibly inflated): quality of supplementary readings, r=.54, and of textbook and lectures, r=.42 for each; use of clarifying examples or illustrations, r=.49; raising of challenging questions/ problems for discussion, r=.40; and freedom to ask questions/ express opinions, r=.38 (p. 334). For the other 10 dimensions, peer-student correlations were low.

As in the Howard et al. study (1985), peers had observational data on which to base their ratings, yet their overall judgments were not as good as those from studies in which raters used their impressions. When rendering judgments on a few specific dimensions of teaching effectiveness (i. e., quality of the textbook used), Centra's (1975) peer raters fared a bit better. Oddly, Centra sugggested dimensions that "colleagues would seem to be able to judge" (p. 335) that either showed poor student-peer correspondence (i.e., quality of exams, r= .17) or which he did not investigate--course syllabi and objectives, and the instructor's qualifications and knowledge.

Both the Aleamoni and Yimer (1973) and Wood (1978) studies appear similiar to studies in which peers received no information on the faculty that they were to rate, and in which peer-student agreement was high; but the peer-student agreement in their studies was low. Procedural differences in these two studies provide some insight into why peer-student correlations were low.

Aleamoni and Yimer asked faculty at one campus to select "three faculty members whom they felt deserved mention for good teaching" (1973, p. 274). The researchers then rated the faculty from 1 to 26 based on frequency of nominations. They collected student ratings on the nominated faculty, from two separate questionaires--the Illinois Course Evaluation Questionaire (CEQ) and the Advisor. Both questionaires used a 4-point rating scale.

The peer-student correlation of .28 from Aleamoni and Yimer's (1973) study was an average of correlations between (a) peer-

student ratings on the Instructor Scale (Advisor) and (b) peer-student ratings on the CEQ Total Scale. Aleamoni and Yimer also reported correlations between peer-student ratings on dimensions of the CEQ: general attitude toward the course, method of instruction, course content, and student interest and attention; but they were all low. The constricted faculty sample (only those who were nominated were rated), and the fact that peers did not actually rate the faculty that they nominated, are possible reasons for the low correlations.

I was unable to locate Wood's (1978) paper. Feldman (1989) reported, however, that he based his correlation of .36 on the mean of correlations across three studies of peer-student correlations--r= .64, .33, and .28, respectively. The highest correlation was based on the correlation of overall rating of faculty by peer raters and the "average score on the three most general questions of the Student Description of Teaching form" (p. 183). The lower correlations were based on the correlations of student ratings on general questions with faculty ranked as the "seven most effective teachers." In these latter two studies, the restriction of the pool of rated faculty may have reduced the size of the correlation between student and peer raters.

Studies with low current student-peer correlations: Overall impressions. Correlations between student and peer ratings of teaching are low when: (a) peers observe faculty and (b) researchers compare only the most effective faculty with student ratings. Correlations between peer-student ratings on specific

dimensions (i.e., students' freedom to ask questions) vary with the dimension being rated and with whether peers have observed the faculty that they are rating. Further, Howard et al. (1985) have demonstrated that variations across validation methods produce different impressions of the validity of any measure of teaching.

Studies with Moderate Current Student-Peer Correlations

Blackburn and Clark (1975), r= .62; Doyle and Crichton (1978), r= .56;Guthrie (1954), p. 51; Bendig (1953), r= .49; Stavridis (1972), r= .46. In the Blackburn and Clark study, 85% of full-time faculty in a small college rated their peers' teaching on a 5-point scale. Students also rated faculty teaching. The correlation of .62 in this study between peer-student teaching ratings occurred when peers did not receive specific information, a condition likely to produce high ratings. The reason that the correlation was not higher is unclear.

Peer raters in Doyle and Crichton's (1978) study were "mostly advanced graduate students," who were more acquainted than typical faculty (p. 816). Students rated one class per instructor. Student and peer raters used a common rating form, consisting of four specific items--"clearly presented subject matter," "was approachable," "got students interested," and "raised challenging questions"--and two general items--overall teaching ability and amount learned (p. 816)

Doyle and Crichton (1978) asked peer raters to "rate colleagues' probable classroom presentation" based on "behavior at faculty meetings, colloquia, and social gatherings" (p. 816). The

correlation between peer-student ratings was .57. It is possible that this correlation was not higher because peer raters used specific, non-classroom based knowledge for their ratings. Peer-student correlations on two of four specific dimensions were moderate (.43)--"raised challenging questions" and "student learning"--attesting to use of specific knowledge. As in the Centra (1975) study, peers showed moderate agreement with students on a few specific dimensions. Generalizing from the present study, however, is difficult, given the atypicality of graduate students as peers.

I could not locate Guthrie's (1954) papers from which Feldman (1989) calculated the .54 peer-student correlation. According to Feldman, some of the student-peer correlations were based on global ratings of teacher effectiveness (p. 179). Unfortunately, the details on these ratings are sketchy. If, however, the mean correlation between peer-student ratings was based on global rating, one would expect this correlation to be higher.

Bendig (1953), r= .49. Students in introductory psychology courses rated 10 instructors at the end of one semester from 1 to 5 on a fourteen-item scale, which included "organization of course material, friendliness toward the students, personal appearance, etc." (p. 333). Using factor analysis, Bendig condensed the items into three major dimensions, writing both positive and negative faculty descriptions for each factor.

Four faculty members "who [knew] the instructors quite well" ranked them from 1 to 5 (Bendig,1953). The median correlation

between peer-student raters across three dimensions was .49. The descriptive information that peers received may have detracted from the correlations between peer-student ratings. Further, using a small number of peer raters and peer-student uses of different rating methods may have reduced the correlations.

In Stavridis' study (1972), students and peer raters who served on departmental promotion committees rated the instructors on an 11-item rating scale, which contained two overall teaching ratings. The correlation for overall ratings was .46. Only two dimensions showed moderate peer-student correlations--instructor's knowledge (r=.38) and arousal of interest in the subject (r=.41).

Because they were on promotion/tenure committees, these peer raters had more information than peer raters in studies reporting high peer-student correlations. In fact, Stavridis (1972) suggested peer-student correlations may have been lower than one might expect due to a correlation (r=.51) between teaching ratings and publications (p. 93).

Studies with moderate student-peer correlations: Overall impressions. Studies in this section required peer raters to use mostly general information for their ratings. One factor that may have reduced the size of the correlations in this group was greater peer familiarity with faculty. Other factors may have been providing descriptive information (Bendig, 1953) and asking subjects to use information from varied settings about faculty (Doyle and Crichton, 1978).

Research Not Included in Feldman's (1989) Meta-analysis

Centra (1994), r=.33; -.03; Kremer (1990), r=.57. In Centra's (1994) study, peers based their ratings on teaching portfolios, which were organized on three dimensions (motivational, interpersonal, and intellectual skills) under which there were 13 items. Student evaluators used the Student Instructional Report scale, (SIR). Centra analyzed peer-student correlations on three SIR scales (organization/ planning, faculty/student interaction, and communications), which best corresponded to portfolios. Two peers rated faculty--one selected by the dean and one selected by each faculty member.

Centra (1994) then investigated correlations between student-peer raters. He found a significant (but low, r=.33) overall correlation between dean-selected peers' and students' ratings. The correlation between faculty-selected peers' and students' ratings was even lower (r=-.03). He concluded that"basing evaluations on a portfolio, particularly for summative purposes, apparently introduces other sources of error" (p.568). He also demonstrated that which peer evaluates one's teaching may be crucial. Again, one can see that peers having access to more information about teaching (even in a teaching portfolio) does not inevitably increase peer-student agreement. In the present study, the correlations between specific dimensions (motivational, interpersonal, and intellectual) for peer-student ratings were no higher than those for overall ratings. In other studies in which peers based their ratings on specific information (for example,

Centra, 1975), correlations between student-peer ratings for certain dimensions were at least moderate.

Kremer (1990) measured the teaching of full-time faculty through student and peer ratings and number of teaching accomplishments (awards, grants related to teaching, and teaching publications). Students used the Purdue CAFETERIA form. He selected peers according to their knowledge of evaluation criteria and of personnel decisions. He told them to use "perception only," not to "look up any information" on the faculty, and to rate their "overall contribution" on a 5-point scale (p. 214).

Kremer (1990) found significant correlations among measures of teaching for peer and student ratings, and for teaching awards. He then averaged the correlations between peer-student ratings (r=.57) and peer ratings-teaching awards (r=.44) and reported a mean correlation of .50 for peer ratings with all teaching measures. He also averaged the .57 correlation between student-peer ratings and the .21 correlation between student ratings-teaching awards, reporting a mean correlation of .39 for student ratings. Finally, he combined the correlations between teaching awards-peer ratings (.44) and between teaching awards-student ratings (.21) and reported a mean correlation of .32 for teaching awards. Using these averaged correlation coefficients, Kremer reported good convergent validity for peer ratings and "adequate" convergent validity for student ratings and teaching awards (p. 215).

He noted that measures of teaching should have lower correlations with research and service than they do with one another (e.g., they should show discriminant validity). He found that the teaching awards measure demonstrated inadequate discriminant validity because its correlation with service (r= .33) was higher than its correlation with the other teaching measures (r=.32). Kremer's reported "good" convergent validity for peer ratings of teaching was based in part on correlation with teaching awards (r=.44), which had inadequate discriminant validity. Kremer's reported "adequate" convergent validity for student ratings was due to their low correlation with teaching awards (r=.21), to which they should not relate.

Kremer (1990) further investigated the discriminant validity of peer evaluations of teaching using one other more "stringent" method (p. 215). He stated that the average correlation of peer ratings (with other teaching measures) should be higher than the average correlations of peer ratings of teaching, research, and service. The average correlation of the peer ratings of teaching, however, was lower than the average correlation for peer ratings of teaching, research, and service combined. Peer ratings of teaching, therefore, showed inadequate discriminant validity.

One may recall that Stavridis (1972) also reported a moderate correlation between peer ratings of teaching and research, which may have partly accounted for a lowered correlation between peer and student ratings. Peer ratings of teaching may not be easily separable from their ratings of service and research. Howard et

al. (1985) cautioned that certain studies of student evaluations may have underestimated their validity due to their having been validated against criteria of lesser validity. Kremer's (1990) study provides some backing for this claim, as both peer evaluations of teaching (under one method) and teaching awards showed inadequate discriminant validity. In common with earlier studies, though, Kremer's peer ratings, based on global, impressions, showed a moderate overall relationship to student ratings (r= .57).

On Broadening the Scope of Peer Review of Teaching

Marsh (as cited in Koon and Murray, 1995) spoke to the ready acceptance of the adequacy of peer review of teaching:

> It is ironic that researchers who argue that the validity of student ratings has not been sufficiently demonstrated, despite the preponderance of research supporting their validity, are so willing to accept other indicators which have not been tested or have been shown to have little validity. (1987, p. 302)

Koon and Murray (1995) refer to such acceptance as "professional-political values" dominating "professional-truth-seeking values" (p. 63).

Feldman (1989, p. 164) cautioned about making assumptions about the validity of varied raters' assessments of teaching, as measured through correlations between raters. He noted that high agreement may provide evidence of interdependent raters--not validity. Low agreement may indicate that although students rate

classroom skills, peers rate "generalized skills" outside the classroom. But Feldman has hit on the central problem: On exactly what "generalized teaching skills" might our peers rate us? Are such generalizations appropriate and fair? Are their contributions unique; or do they (at best) overlap with student judgments?

Although peer evaluation of faculty teaching, or "taking teaching seriously," (Seldin, 1993, p. 1) has enjoyed a surge of interest, it is not safe to assume peer evaluating is more effective than student evaluating--even if peers base their evaluations on observations or portfolios. Ironically, many of the studies that this article reviewed point to the conclusion that faculty might be better-served by peers' general impressions of teaching, which, in turn, may have been informed by students. At best, there is meager information on specific dimensions of teaching that peers judge well, even if one provides them with a well-organized teaching portfolio. One must conclude that the areas of teaching that our peers can capably judge is yet to be determined.

There is, however, untapped potential for peers to aid in evaluating teaching. Marsh (1993) eloquently demonstrated how peers might aid in interpreting profiles of faculty performance gleaned from almost 1 million student evaluations from a 13-year period. His data revealed that individual faculty showed distinct profiles on the Students' Evaluation of Educational Quality instrument--on dimensions such as organization, enthusiasm, student's perceived learning, and so forth (p. 8). Theall and

Franklin (1991) also wrote a richly-informative chapter on how to use the results of student evaluations to improve teaching.

Despite the promise of increased peer involvement in teaching evaluation, the evidence for peers' effectiveness in broadened evaluative roles is scant and inconsistent. Although political pressures to elevate the status of teaching are great, one must qualify assertions, such as peer evaluation of teaching is the "right thing to do" (Hutchings, 1996).

References

Aleamoni, L. M., & 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.

Ballard, M., Rearden, J. & Nelson, L. (1976). Student and peer rating of faculty. Teaching of Psychology, 3, 88-90.

Bendig, A. W. (1953). An inverted factor analysis study of student-rated introductory psychology instructors. Journal of Experimental Education, 21, 333-336.

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

Cashin, W. E. (1989). Defining and evaluating college teaching. (Idea Paper No. 21). Manhattan, Kansas: Kansas State University, Center for Faculty Evaluation and Development.

Centra, J. A. (1975). Colleagues as raters of classroom instruction. Journal of Higher Education, 46, 327-337.

Choy, C. (1969). The relationship of college teacher effectiveness to conceptual systems orientation and perceptual orientation. Unpublished doctoral dissertation, Colorado State College, Greeley.

Centra, J. A. (1994). The use of the teaching portfolio and student evaluations for summative evaluation. Journal of Higher Education, 65, 555-570.

Cohen, P. A., & McKeachie, W. J. (1980). The role of colleagues in the evaluation of college teaching. Improving College and University Teaching, 28, 147-154.

Doyle, K. O., & Crichton, L. L. (1978). Student, peer, and self evaluations of college instructors. Journal of Educational Psychology, 70, 815-826.

Feldman, K. A. (1989). Instructional effectiveness of college teachers as judged by teachers, themselves, current and former students, colleagues, administrators, and external (neutral) observers. Research in Higher Education, 30, 137-194.

Howard, G. S., Conway, C. G., & Maxwell, S. E. (1985). Construct validity of college teaching effectiveness. Journal of Educational Psychology, 77, 187-196.

Hutchings, P. (1996). The peer review of teaching: Progress, issues, and prospects. Innovative Higher Education, 20, 221-234.

Koon, J., & Murray, H. G. (1995). Using multiple outcomes to validate student ratings of overall teacher effectiveness. Journal of Higher Education, 66, 61-81.

Kremer, J. F. (1990). Construct validity of multiple measures in teaching, research, and service and reliability of peer ratings. Journal of Educational Psychology, 82, 213-218.

Magner, D. K. (1997, September 5). Report says standards used to evaluate research should also be used for teaching and service. The Chronicle of Higher Education, 44, A18-A19.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions

for future research. International Journal of Educational Research,17, 253-388.

Marsh, H. W., & Bailey, M. (1993). Multidimensional students' evaluations of teaching effectiveness. Journal of Higher Education, 64, 1-17.

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

McCarbery, R. J. (1970). The relationship of selected factors to ratings of community college teachers. Unpublished doctoral dissertation, Indiana University, Bloomington.

McMillan, J. H. (1996). Educational research: Fundamentals for the consumer. New York: Harper Collins College Publishers.

Scriven, M. (1995). Student ratings offer useful input to teacher evaluations. Washington, DC: ERIC Clearinghouse on Assessment and Evaluation. (ERIC Document Reproduction Service No. ED398240)

Seldin, P. (1993). Successful use of teaching portfolios. Bolton, MA: Anker Publishing Co.

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, Tempe.

Theall, M., & Franklin, J. (1991). Using student ratings for teaching improvement. In M. Theall & J. Franklin (Eds.),

EffectivePractices for Improving Teaching: No. 48. New Directions for Teaching and Learning. San Francisco: Jossey-Bass, Inc.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

Reproduction Release

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Peer Evaluation of Teaching: Claims vs. Research	
Author(s): Candace W. Burns	
Corporate Source: University of Arkansas at Little Rock	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, Resources in Education (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign in the indicated space following.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker sh
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSIC DISSEMIN. MICROFICHE OI TO THE EDU INFORM.
Level 1	**Level 2A**	
X		
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g. electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level dissemir

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche, or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: Candace W. Burns	Printed Name/Position/Title: Candace W. Burns, Associate Professor	
Organization/Address: Dept. Of Educational Leadership University of Arkansas at Little Rock 2801 S. University Little Rock, AR 72204	Telephone: (501) 569-3267	Fax: (501) 569-8694
	E-mail Address: cwburns@ualr. edu	Date: June 18, 1998

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598
Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 9/97)