RE-EVALUATING STUDENT EVALUATION OF TEACHING: THE TEACHING EVALUATION FORM

Terry A. Wolfer
University of South Carolina

Miriam McNown Johnson
University of South Carolina

This study reports on the aggregate analysis of scores generated by a standardized instrument, the Teaching Evaluation Form (TEF; Hudson, 1982), at the College of Social Work, University of South Carolina. The data included more than 11,000 completions of the instrument in 508 class sections offered during a 4-year period. The analysis revealed a severely negatively skewed and truncated distribution of scores, with no low outlying instructors. It raises questions about the TEF's usefulness for either administrative or teaching improvement purposes. In light of these questions, the paper discusses and recommends alternate approaches to evaluation of teaching in social work education.

CONTROVERSY CONTINUES regarding the validity of student evaluations of teaching, with evidence for and against their use as measures of instructor teaching performance (Abrami, d'Apollonia, & Cohen, 1990; Cashin & Downey, 1992; Dwinell & Higbec, 1993; Hepworth & Oviatt, 1985; Greenwald, 1997; McKeachie, 1997; Solas, 1990). Some studies identify various factors that may influence student assessments of instructor performance (Marsh, 1984; Petchers & Chow, 1988; Weinbach, 1988) including class size (Hanna, Hoyt, & Aubrecht, 1983), course content (Cashin, 1990; Hanna et al., 1983), gender of the instructor (Anderson & Miller, 1997; Martin, 1984), and grading leniency (Greenwald & Gillmore, 1997).

Perhaps because of the controversy, schools have tried various methods for obtaining student evaluations, ranging from semi-structured, qualitative measures to standardized, exclusively quantitative measures. Furthermore, some schools have experimented with other methods for evaluating teaching performance such as instructor self-evaluation, review of course materials and teaching portfolios compiled by instructors, and in-class observation by faculty peers. Nevertheless, student evaluations remain the most widely used method for evaluating teaching. Students remain the most common source of data and standardized measures the most common method for soliciting these data (Johnson & Wolfer, 2001).

Journal of Social Work Education Vol. 39, No. 1 (Winter 2003).

Purpose of Student Evaluation of Teaching

However little confidence instructors place in student evaluations, they continue to be widely used in higher education, including social work education. Whether or not it is completed by students, evaluations of teaching have two primary purposes: administrative decision making and teaching improvement (McKeachie, 1997; Pike, 1998). As Marsh (1987, 1991) has suggested, these two purposes correspond with summative and formative types of evaluation. Looked at this way, Cashin and Downey (1992) suggest that summative evaluation "focuses on using student ratings to make a final judgment about an instructor's teaching effectiveness," while "formative evaluation involves using the ratings diagnostically, that is, to make a decision about possible ways to improve teaching" (p. 563). Higher education administrators and tenure committees continue to depend on student evaluations of teaching for making decisions about instructor hiring, promotion, tenure, salary adjustment, and retention. As a result, "Student ratings can make or break the careers of instructors on grounds unrelated to objective measures of student learning" (Ceci & Williams, cited in Wilson, 1998).

What is less apparent, however, is the extent to which instructors actually use teaching evaluation data to improve their teaching performance. Periodic review of evaluation data by administrators and tenure committees certainly may provide an incentive for instructors to pay attention to these data but there is little, if any, research in social work education about the use of these data by individual instructors to improve their teaching performance or the appropriateness of available data for this purpose.

Different Information Needed for Different Purposes

As two authors suggest, it is essential that one is clear about the purpose of evaluating teaching and thereby select methods best suited for particular purposes (McKeachie, 1997; Pike, 1998). While the two purposes mentioned above are closely related, they may require somewhat different types of data. For administrative purposes it will be important to have an overall summary of teaching ability while efforts to improve teaching will benefit most from data that discriminate quite specifically between a teacher's areas of strength and weakness.

Furthermore, these two purposes suggest the need for different information when interpreting teaching evaluation data. For administrative purposes, we need to know whether any contextual factors (e.g., instructor demographics, course characteristics) influence student evaluations of teaching and should be considered in making fair administrative decisions (e.g., Jirovec, Ramanathan, & Alvarez, 1998; Petchers & Chow, 1988; Pike, 1998). For teaching improvement purposes, we need to know whether an instrument adequately identifies and distinguishes problematic aspects of teaching performance so these can be specifically addressed (i.e., what patterns appear; McKeachie, 1997; Pratt & Associates, 1998).

Re-evaluating Teaching Evaluation Data

These divergent purposes suggest several research questions:

1. How are student evaluation of teaching scores (both item and total means) distributed for all instructors in general, and for low-scoring instructors in particular (i.e., those one or more than one *SD* below the overall mean)?
2. What, if any, factors (e.g., instructor or course variables) influence student evaluation of teaching scores?
3. What, if any, patterns exist in individual scale items for low-scoring instructors (i.e., those one or more than one *SD* below the overall mean)?

Method

To assess the suitability of one teaching evaluation instrument, we re-evaluated data generated by its use at the College of Social Work at the University of South Carolina over a period of 4 years. We re-evaluated data previously examined by individual instructors and previously used by tenure and promotion committees and the dean.

This project was carried out in a large public university with a racially diverse student body. The university offers both a Master's of Social Work (MSW) and a doctoral (PhD) degree. At the time of the study, the MSW program offered two concentrations in the advanced year: micro practice (with individuals, families, and groups) and macro practice (with organizations and communities). The program had a total enrollment of approximately 500 MSW students, including about 50 advanced-standing students and 150 part-time students. At any one time, the faculty comprised about two dozen full-time, tenure-track instructors. They were equally divided by gender, although men were disproportionately represented in the highest rank. Adjunct instructors and PhD students taught some courses.

At the end of every term, students evaluated their instructors using the Teacher Evaluation Form (TEF) developed by Hudson (1982). On this instrument, students respond to 26 items using a 10-point scale anchored by letter grades, with 0="F," 2="D," 4="C," 6="B," 8="A" and 9="A+." Items cover a broad range of "qualities which describe instructor performance" such as ability to create interest in the material, apparent preparation for the course, availability to students outside of normal class hours, and use of clear evaluation standards. The TEF is reported to have content and construct validity, and an "extraordinarily high" level of internal consistency (Pike, 1998, p. 269).

We received formal permission from our colleagues to have access to the evaluation results. They agreed on the condition that information that could be used to identify individuals—such as instructor codes and course section numbers—would be viewed only by statisticians in another university department. Following our instructions, statisticians replaced individual instructor identifiers with new codes (e.g., gender, rank), generated descriptive statistics, and analyzed relationships between TEF scores and characteristics of instructors and courses.

The sample comprised more than 11,000 ratings from 508 class sections of MSW courses

offered by the program from fall, 1994, through summer, 1998. The selected time period began with the college-wide adoption of the TEF and ended immediately before a series of curricular revisions that altered course content and sequencing. Broken down by curricular area, there were 43 (8.6%) human behavior sections, 37 (7.4%) policy sections, 92 (18.5%) research sections, 129 (25.9%) foundation practice sections, 93 (18.7%) advanced micro practice sections, 48 (9.6%) advanced macro practice sections, 56 (11.2%) elective sections, and 10 unclassified sections. More than three quarters (n=352, 77.9%) of the sections were taught using a traditional classroom format, with the others (n=100, 22.1%) being offered through interactive distance education or missing (n=56). All distance-education sections were non-practice foundation year courses or electives. About one fifth (n=106, 20.9%) of the sections were offered during summer terms, with the remainder split between the fall (n=197, 38.8%) and spring (n=205, 40.4%) terms.

In addition to course variables, instructor-characteristic codes linked to each section included gender, employment status, and teaching experience. To ensure anonymity, rank was used as a proxy for years of experience, and race and ethnicity were not coded. Accurately reflecting the gender breakdown of the school's faculty, men and women each taught about half of the sections (n=265 [52%] and n=243 [48%], respectively). Full-time, tenure-track instructors taught about 80% (n=399) of the sections: assistant professors (23.6%, n=118), associate professors (21.8%, n=108), and full professors (34.3%, n=172). Adjunct instructors and PhD students taught the remaining 20% (n=99) of the sections.

Results

Item-by-item analysis revealed little variation. Means for the 26 items across all sections ranged from a high of 8.24 on item 2 ("apparent knowledge of course material") to a low of 7.65 on item 6 ("performance as a lecturer"). As a result, all item means fell in the "A" range of the scale. The greatest variation on any item (SD=1.01) was also on item 6.

Because of the lack of variability across items in the present analysis, and because previous evaluation (Pike, 1998) of the TEF instrument indicated that it is a one-factor or unidimensional scale, further analyses were conducted using only the mean for all 26 items for each class section. The mean of all items across all sections was 7.92 (SD=0.77) and the median was 8.1. Readers are reminded that this was a 10-point scale with the top ratings being 8 ("A") and 9 ("A+").

In addition, the amount of variability in the distribution of TEF scores was very limited (see the histogram in Figure 1). For half of the sections, instructors received mean scores that fell above 8.11 (between an "A" and an "A+"), reflecting a seriously skewed and truncated distribution. The skewness was -1.58 and the standard error was 0.034. A skewness value more than twice the standard error indicates a significant departure from symmetry.

We analyzed results by several course characteristics. Ratings did not vary significantly by year (1994–1998), by term (fall, spring, or summer), or by format (on-campus or distance education). Comparison of results for sections grouped by curricular area did reveal interesting patterns. Instructors teaching sections of elective courses received the

highest ratings (M=8.18, SD=0.64), followed by advanced micro practice (M=8.02, SD=0.66), advanced macro practice (M=7.96, SD=0.85), policy (M=7.96, SD=0.64) and foundation practice (M=7.87, SD=0.71). Instructors teaching HBSE (M=7.66, SD=0.72) and research (M=7.62, SD=0.93) were rated lowest. However, the differences were not statistically significant.

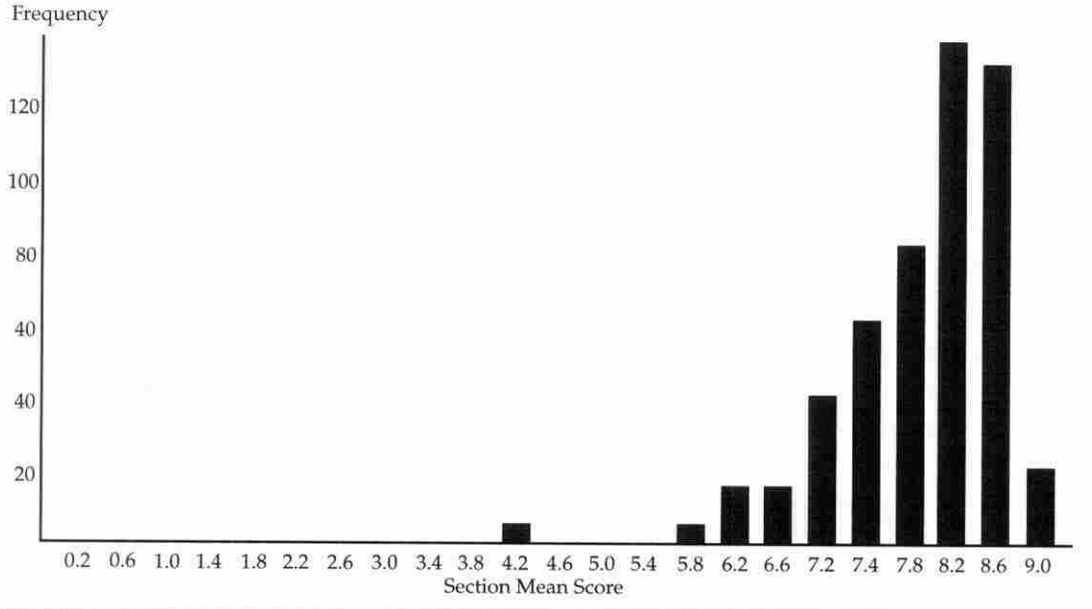
Next, we compared results by instructor characteristics. Evaluations for sections taught by adjunct instructors were slightly higher (M=8.02, SD=0.65) than were evaluations for tenure-track instructors (M=7.93, SD=0.76) and PhD students (M=7.77, SD=0.81), but again, the differences were not statistically significant. For tenure-track instructors, there was an inverse correlation between TEF scores and experience (rank): the average rating for sections taught by full professors was 7.84 (SD=0.84), for associate professors it was 7.95 (SD=0.72), and for assistant professors it was 8.04 (SD=0.67). But, once again, the results were not significant. However, an independent t test showed that sections taught by women instructors received significantly higher evaluations (M=8.11, SD=0.65) than those taught by men (M=7.74, SD=0.83; $p<.0001$).

As indicated by research question 3 on p. 113 above, we planned further analyses of low outliers. However, although a few instructors received low mean scores on the TEF for particular class sections, no instructor earned an overall personal mean score one standard deviation or more below the overall college mean.

Discussion

The most serious limitation of the study was the selection of a convenience sample. Although more than 500 course sections were

FIGURE 1. Distribution of TEF Means of All Items Across All Sections, by Class Section (*N*=508)



evaluated, all were taught within the same program by a small number of individuals (N=67). In addition to possible commonalities among social work instructors and students, consistent hiring and retention patterns within the program and informal university and college norms for instructor evaluation may result in considerable homogeneity of ratings for instructors in one program. These factors may partly explain the results of this study and therefore limit their generalizability to other programs. Furthermore, we used no other sources of evaluation data (e.g., peer or administrative reviews, students' comments, self-assessments by instructors) that might have supported the validity of the TEF evaluations.

Although not statistically significant, differences between course sections across the curriculum were consistent with findings in other studies (Jirovec et al., 1998; Petchers & Chow, 1988). We note that students taking electives and advanced practice courses (those that received the highest ratings in the study) have exercised some degree of choice in selecting them, whereas policy, foundation practice, HBSE, and research courses are all required. It is also true that enrollment in practice sections in this program is, by design, more limited in size than sections of other courses. Studies consistently show an association between smaller class size and higher student ratings (e.g., Hanna et al., 1983).

Several researchers suggest that students rate instructors of the same gender higher (Anderson & Miller, 1997; Martin, 1984). This might explain why, with a student body that is overwhelmingly female (approximately 90%), female instructors receive higher TEF scores.

It was disconcerting to note that ratings did not show a positive correlation with experience (rank). Additional analysis revealed that results were confounded by gender differences. Unfortunately, a breakdown of the data set by a combination of gender and rank made the findings for some categories dependent on the performance of only two or three individuals so discussion is limited both by sample size and by sensitivity to issues of confidentiality.

The findings were consistent with the literature comparing global and multidimensional evaluation instruments (e.g., Cashin & Downey, 1992). It appeared that the very limited range of actual scores for the TEF makes it difficult to identify meaningful distinctions among instructors. The fact that no instructors scored one standard deviation or more below the college mean supports this conclusion. In addition, the unidimensional character of the instrument makes it difficult to identify patterns of performance deficits that could be used to guide efforts toward individual improvement or planning for department-wide training.

Implications

TEF scores in this study were very high and showed little variation. We note, however, that our college mean actually fell below those previously reported for three of five schools (Pike, 1998). In other words, the scores were more highly skewed and showed less variability for three schools than in the present study. As a result, the TEF's measurement limitations may be even more problematic than suggested by the current study alone. This becomes more evident when one reconsiders the two primary purposes for teaching evaluation.

As noted in the introduction, teaching evaluation has two primary purposes in social work education: administrative decision-making (summative) and teaching improvement (formative). Failing to distinguish these dual purposes may limit the usefulness of any measurement tool or strategy. We conclude that the TEF does not address either of these two purposes well and offer several recommendations for alternative, more targeted measurement strategies.

Administrative Decision Making

Some education researchers have long maintained that criterion-based measures, rather than peer comparisons, should be used for evaluating teaching. "Without a set of norms [or criteria] . . . it is difficult to interpret student ratings. Lack of adequate norms is not so much a flaw in the studies, which represents standard practice fairly, as it is a failure in standard practice itself" (Kulik & McKeachie, 1975, p. 225). For this reason, it is important that researchers develop and administrators adopt criterion-based measures.

Following d'Apollonia and Abrami (1997), we recommend that student evaluations be used to make "only crude judgments of instructional effectiveness (exceptional, adequate, and unacceptable)" (p. 1205). Ample research supports the validity of student evaluations for making these relatively crude distinctions among instructors but casts doubt upon their usefulness for making finer distinctions. Unfortunately, the apparent precision of numerical scores generated by the TEF may mistakenly imply a level of measurement precision that simply does not exist. As McKeachie (1997) suggests, the "presentation of numerical means or medians (often to two decimal places) leads to making decisions based on small numerical differences—differences that are unlikely to distinguish between competent and incompetent teachers" (p. 1223).

Furthermore, "the use of norms not only leads to comparisons that are invalid but also is damaging to the motivation of the 50% of faculty members who find that they are below average" (McKeachie, 1997, p. 1223). Rather than base administrative decisions on between-instructor comparisons, colleges may be better served by comparing individual instructor performance with an objective standard. For example, administrators could establish a cutoff point below which teaching effectiveness is considered inadequate. Alternately, such a cutoff might be established at one or two standard deviations below the mean.

Consistent with the literature on student evaluation of teaching, results from this study suggest that some consideration may be given to weighting student responses in different types of courses. For example, consideration may be given to curricular area, the degree of choice the students have in selecting the course, and typical class size.

More generally, however, these results suggest caution when interpreting minor variations in statistical data. Efforts to account for such variations across curricular areas, for example, may lead to making distinctions between instructors that are essentially meaningless. Instead, it may be better to use more simple, global evaluations by students for purposes of personnel evaluation.

Based on their research, Cashin and Downey (1992) recommend using a single global measure of teaching effectiveness such as "Overall, I rate this INSTRUCTOR an excellent teacher" (p. 569). For purposes of retention and promotion, administrators and tenure committees need only determine that an instructor generates some minimum percentage of satisfactory or excellent ratings or does not exceed some maximum percentage of unsatisfactory ratings. Greater measurement precision seems unwarranted and unnecessary for administrative purposes.

Feedback and Improvement

At the same time, instructors need more and different information about how to improve their teaching, and they need it earlier in order to make necessary adjustments. Too often evaluation occurs only at the end of a term when it is too late to remedy classroom problems, or otherwise contribute to the learning of current students. When administered at term's end, any instrument can only provide information for future teaching adjustments. But as McKeachie (1997) noted, such adjustments may not be appropriate for subsequent classes.

Regardless of when it is used, the TEF does not provide the type of data necessary for making teaching adjustments. Because there is little between-item variation, the instrument provides minimal direction for individual instructors on how to improve. Because the scale is unidimensional, it provides no clustering of behaviors that warrant remediation.

Furthermore, because it focuses on discrete instructor behaviors, the TEF diverts attention from the fundamental interaction in classrooms between instructors and students. McKeachie (1997) notes,

> Effective teaching is not just a matter of finding a method that works well and using it consistently. Rather, teaching is an interactive process between the students and the teacher. Good teaching involves building bridges between what is in your head and what is in the students' heads. What works for one student or for one class may not work for others. (p. 1224)

The TEF implies that particular teaching behaviors have value in and of themselves. We suggest, instead, that teaching behaviors can only be understood in the relational context of particular classrooms. In other words, a specific teaching behavior may be present and highly regarded in one classroom or present but less well received in another. It is not the teaching behavior per se but how this behavior fits within the particular group of students that is of greatest importance. From this perspective, instructors will benefit from more detailed and timely written or oral student feedback about what is or is not working than that provided by the TEF (or other end-of-term standardized evaluation instruments).

In addition, learning to attend to and take account of classroom dynamics in an ongoing way may be an important strategy for instructors. Educators must give up notions of teaching "excellence" in the abstract (i.e., as operationalized by scores on standardized instruments like the TEF), and realize instead

that excellent teaching occurs in particular classrooms. And excellent teaching depends in large measure on the instructor's ability to respond to the particular students in those classrooms. That ability to learn and respond to students is related to student learning in the classroom. But it is not a *measure* of student learning. For that reason, we do not recommend merely evaluating relational dynamics instead of teaching behaviors.

As Jirovec et al. (1998) report, students appreciate instructors who are well organized, have good rapport with students, and use clear and fair grading procedures. Although it seems reasonable that these instructor behaviors are associated with effective teaching, they are not synonymous with student learning. In other words, they may facilitate but do not indicate student learning. Put differently, student ratings of teaching behaviors are more a measure of student satisfaction than of student learning, the ultimate criteria of excellent teaching. For that reason, it seems to us that student-learning outcomes warrant relatively greater attention from instructors and administrators than teaching behaviors.

Our evaluation of TEF data led to this further insight: the TEF may be better understood as a measure of consumer satisfaction with teaching behaviors than of learning outcomes. While many of the teaching behaviors identified in the TEF have been correlated with teaching effectiveness, the scale measures the presence of these behaviors rather than the learning outcomes of a teacher's efforts. Such learning outcomes are analogous to the outcomes of direct practice, and the real measure of teaching effectiveness.

For evaluating teaching effectiveness, McKeachie (1997) prefers "student ratings of attainment of educational goals rather than [ratings of teaching behaviors]" (p. 1218). As he explains,

> Many students prefer teaching that enables them to listen passively—teaching that organizes the subject matter for them and that prepares them well for tests. . . . Cognitive and motivational research, however, points to better retention, thinking, and motivational effects when students are more actively involved in talking, writing, and doing. (p. 1219)

No doubt, it is difficult to accurately measure student learning, especially in professional courses where the goals go beyond mere knowing to include doing. Nevertheless, this difficulty does not relieve individual instructors or their programs of responsibility for measuring student outcomes. Furthermore, engaging with students in the process of evaluating their learning can promote self-reflection for both instructors and students and can provide important feedback on the immediate learning process. Indeed, "student ratings of their attainment of educational objectives not only provide better data for personnel committees but also stimulate both students and teachers to think about their objectives—something that is educational itself" (McKeachie, 1997, p. 1223). Such a shift in focus may serve to improve teaching and learning more efficiently than a continued focus on teaching behaviors.

Perhaps our emerging perspective can be clarified by analogy. The recent and ongoing

push for evaluation of social work practice in agencies focuses on the outcomes of intervention in the lives of clients. It emphasizes changes resulting from intervention and pays little attention to specific social worker behaviors to promote those changes. And while researchers continue to be (appropriately) interested in social worker behaviors (e.g., in developing manual-based interventions), they generally do not rely on client ratings of these behaviors (and they do not rely exclusively on client ratings of social worker behaviors). In a direct practice setting, it would strike people as odd and inappropriate to base a social worker's pay or promotion on client ratings of the social worker's behaviors, especially in the absence of attention to client outcomes. Of course, measuring their clients' outcomes may produce some anxiety for social workers just as measuring students' learning outcomes may for instructors. It seems ironic that social work instructors might teach their students to measure client outcomes but not evaluate their own teaching practice in light of student-learning outcomes.

Conclusion

In summary, the authors conclude that the information gained from the TEF does not warrant the effort required to administer it. On the one hand, less information may actually serve the purpose of making personnel decisions. On the other hand, more and different information may be required for actually improving teaching, and that information will be most useful before the end of the term.

Consistent with current trends, we believe individual social work instructors, tenure and promotion committees, and administrators must evaluate instructors' teaching performance. However, based on our re-analysis of TEF data, we encourage educators to clearly specify their evaluative purpose and select a measurement strategy consistent with that purpose. For summative evaluation, for example, we recommend focusing on student-learning outcomes and including student self-assessment as one part of the process. For formative evaluation, we recommend collecting information during the term rather than at the end and broadening the focus to include the classroom environment.

References

Abrami, P. C., d'Apollonia, S., & Cohen, P. A. (1990). Validity of student ratings of instruction: What we know and what we do not. *Journal of Educational Psychology, 82*, 219–231.

Anderson, K., & Miller. E. D. (1997). Gender and student evaluations of teaching. *PS: Political Science & Politics, 30*, 216–219.

Cashin, W. E. (1990). Students do rate different academic fields differently. *New Directions for Teaching and Learning, 43*, 113–132.

Cashin, W. E., & Downey, R. G. (1992). Using global students rating items for summative evaluation. *Journal of Educational Psychology, 84*, 563–572.

Dwinell, P. L., & Higbec, J. L. (1993). Students' perceptions of the value of teaching evaluations. *Perceptual and Motor Skills, 76*, 995–1000.

d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist, 52*, 1198–1208.

Greenwald, A. G. (1997). Validity concerns and usefulness of student rating of instruction. *American Psychologist, 52*, 1182–1186.

Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist, 52,* 1209–1217.

Hanna, G. S., Hoyt, D. P., & Aubrecht, J. D. (1983). Identifying and adjusting for biases in student evaluations of instruction: Implications for validity. *Educational and Psychological Measurement, 43,* 1175–1185.

Hepworth, D., & Oviatt, B. E. (1985). Using student course evaluations: Findings, issues, and recommendations. *Journal of Social Work Education, 21,* 105–112.

Hudson, W. W. (1982). *The clinical measurement package: A field manual.* Homewood, IL: Dorsey.

Jirovec, R. L., Ramanathan, C. S., & Alvarez, A.R. (1998). Course evaluations: What are social work students telling us about teaching effectiveness? *Journal of Social Work Education, 34,* 229–236.

Johnson, M. M., & Wolfer, T. A. (2001). [Methods of evaluating teaching in social work education programs: A national survey]. Unpublished raw data.

Kulik, J. A., & McKeachie, W. J. (1975). The evaluation of teachers in higher education. In F. N. Kerlinger (Ed.), *Review of research in education* (Vol. 3, pp. 210–240). Itasca, IL: Peacock.

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

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11,* 253–388.

Marsh, H. W. (1991). Multidimensional students' evaluations of teaching effectiveness: A test of alternative higher order structures. *Journal of Educational Psychology, 83,* 285–296.

Martin, E. (1984). Power and authority in the classroom: Sexist stereotypes in teaching evaluations. *Journal of Women in Culture and Society, 9,* 482–492.

McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist, 52,* 1218–1225.

Petchers, M. K., & Chow, J. C. (1988). Sources of variation in students' evaluations of instruction in a graduate social work program. *Journal of Social Work Education, 24,* 34-42.

Pike, C. K. (1998). A validation study of an instrument designed to measure teaching effectiveness. *Journal of Social Work Education, 34,* 261–271.

Pratt, D. D., & Associates. (1998). *Five perspectives on teaching in adult and higher education.* Malabar, FL: Krieger.

Solas, J. (1990). Effective teaching as construed by social work students. *Journal of Social Work Education, 26,* 145–154.

Weinbach, R. (1988). Manipulation of student evaluations: No laughing matter. *Journal of Social Work Education, 24,* 27–34.

Wilson, R. (1998, January 16). New research casts doubt on value of student evaluations of professors. *The Chronicle of Higher Education,* p. A12.

Accepted: 10/02.

Terry A. Wolfer is associate professor and **Miriam McNown Johnson** is associate professor, College of Social Work, University of South Carolina.

Address correspondence to Terry A. Wolfer, College of Social Work, University of South Carolina, Columbia, SC 29208; email: terry.wolfer@sc.edu.