University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

POD Quarterly: The Journal of the Professional and Organizational Development Network in Higher Education

Professional and Organizational Development Network in Higher Education

7-1-1980

Letters to the Editor

Follow this and additional works at: http://digitalcommons.unl.edu/podqtrly

Part of the Higher Education Administration Commons

"Letters to the Editor" (1980). *POD Quarterly: The Journal of the Professional and Organizational Development Network in Higher Education.* Paper 65.
http://digitalcommons.unl.edu/podqtrly/65

This Letter to the Editor is brought to you for free and open access by the Professional and Organizational Development Network in Higher Education at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in POD Quarterly: The Journal of the Professional and Organizational Development Network in Higher Education by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Letters to the Editor

To The Editor:

Please excuse our anger! The POD Quarterly *is a refereed journal, is it not? How then was it possible that the article by Stevens and Marquette, "Differing Student and Faculty Perceptions of Teaching Effectiveness and the Value of Student Evaluation," appeared in the Winter, 1979 issue? A couple of years ago we began working with several colleges to help them improve their evaluation systems, partly because there seemed to be such confusion and misinformation abroad. Unfortunately, the Stevens and Marquette article may perpetuate this state of affairs. It is on a par with the infamous Rodin and Rodin article published in* Science *in 1972, one of the few well-known pieces which they cite.*

Fortunately, the article by Peter Seldin published in the Spring, 1980 issue of POD Quarterly, *"Guidelines for Successful Student Evaluation Programs," is a sound and sensible treatment of the topic. However, we cannot restrain ourselves from casting out a few errors here. Stevens and Marquette seem to have based their article on faulty hypotheses, very thin research, numerous myths and reliance on what can be called, charitably, a provincial literature.*

The major false hypothesis is that there are usually differences between student and faculty perceptions of teaching effectiveness. There are not. In the first place, much more thorough scholars (cf., Eble, The Craft of Teaching [*San Francisco, 1976*], *Hildebrand, Wilson and Dienst,* Evaluating University Teaching [*Berkeley, 1971*], *and Kulik and McKeachie, "The Evaluation of Teachers in Higher Education," in Kerlinger,* Review of Research in Education [*Itaska, Illinois, 1975*]*) indicate that there is significant correlation between student and faculty perceptions of what contributes to good teaching. The traits examined by Stevens and Marquette undoubtedly produced mixed reactions because they are a hodgepodge of elements of both teacher behaviors and course environment which are only speciously related to a common definition of effective instruction. More important, their second hypothesis, that there are differing perceptions between faculty and students as to the appropriate use of student ratings, need not have any real effect on the value of such data. There are three common and appropriate uses for student ratings, all of which may be appreciated by both faculty and students if the uses are publicly agreed upon and the instruments appropriately designed. Student*

ratings, used as parts of comprehensive evaluation systems, can be and are effective tools for making personnel decisions, for improving teaching and for providing students with information about instructors and courses. This is not news: such values (and good advice about how to realize them) have been discussed in the literature for more than a decade.

The literature Stevens and Marquette cite is limited almost entirely to the field of business management and decision sciences. In fact, the only reference notes in the article refer to three papers presented at the March, 1978 conference of the Western American Institute for Decision Sciences meeting in San Diego. Now that's provincial! Their bibliography includes only three widely recognized, respectable works on student ratings, the most recent of which is from 1976. A quick check of the ERIC Clearinghouse for Higher Education would have turned up better resources.

The conclusions drawn by Stevens and Marquette are stranger still. They invoke the red herring of federal law enforcement of employee selection procedures to suggest that student evaluations be used exclusively for faculty development (out goes the baby with the bath water). They call for "future areas of research" on self-ratings, peer ratings, student ratings and supervisor ratings, suggesting that classroom visitations (by far the most problematic form of evaluation) become a part of the process —another sign that they are simply unaware of the work going on around the country during the last decade. Thus, it should not be surprising that they end with a confused section on job analysis and student evaluations qua performance ratings by subordinates. Some faculty may view students from such a deplorable perspective; they certainly should not be encouraged to do so in the pages of the POD Quarterly.

Stephen C. Scholl
Joan North

STEVENS AND MARQUETTE REPLY:

Dear Editor:

First, we would like to say that we are pleased that Scholl and North are working with several colleges to help them improve their evaluation systems. The task is so often overlooked. In this reply to their comments, we prefer not to launch personal attacks upon the POD Quarterly, *the* Journal's *reviewers, or its contributors, as Scholl and North have (we believe inappropriately) seen fit to do. Our energies will instead be devoted to responding to the more intelligible of Scholl and North's comments and criticisms. This is not always easy since they have not made clear to us how we have added to "confusion and misinformation abroad." In fact, even their specific criticisms seem, at times, to lack coherence. For example, when they say ". . . we can not restrain ourselves from casting out a few errors here," we are not sure if they mean that they have a few*

inaccurate things to say, or whether they simply believe that they have the right to reprove us for submitting our article and to castigate the reviewers for accepting it. It would seem that these scholars would have been better served by writing to us directly and elaborating, after a period of deliberation, upon the specific points with which they took issue.

We, of course, recognize that the study has its limitations. For example, the teaching traits selected are just a few of those we might have examined. We selected these (A) *to be consistent with a study by Baum and and Brown (1978),* (B) *to parallel the traits discussed by Costin,* et al. *(1971), and* (C) *to maintain a relatively short instrument, since we have discovered from previous studies that the longer the instrument, the less likely that busy faculty members will complete it. As for describing this research as "very thin," we assume that Scholl and North refer to the sample size. We have since administered the instrument to an additional 226 faculty members and 572 students from four other universities. The results reported in the Winter, 1979 issue of* POD Quarterly *do not require revision based on these additional cases. Furthermore, it should be remembered that small sample sizes make the attainment of statistical significance more difficult and that the original sample was considerably larger than samples which would be considered small in a statistical sense.*

We must concur with Scholl and North on one point; we are doubtless not as well read in the higher education literature as they purport to be. This does not mean, however, that we are unaware of this literature, nor is it an admission of being as provincial as these writers claim we are. We built a file of over 50 relevant articles on teaching effectiveness, the majority of which were found in journals of higher education, not business publications. We did not feel it appropriate to reference all these articles, however, and only those most germain to our study were included. Of those referenced, approximately one-third come from journals outside business and decision sciences. The percentage of referenced higher education articles is even higher. Furthrmore, if our bibliography is lacking we can feel secure in the knowledge that famous, more widely respected works such as Scholl and North's Workbook *have been referenced in the* ERIC Clearinghouse. *It should also be noted that those articles listed under the heading "bibliography" in the article are, in fact, referenced in the text. The division of the references into three "reference notes" and 23 "bibliography" items was a decision made by the* Journal's *editorial staff. Nevertheless, since we are both in business administration, we did seek articles on teaching effectiveness in the business literature, not because we are provincial, but because our sample consisted of business students and faculty.*

We are appaled that there are people who neither talk to others outside their disciplines nor share ideas and research findings, and yet feel the need to denigrate the works of those outside their own disciplines (we are not *specifically referring to School and North). To combat this problem at a personal level, we have become active members in several professional*

organizations, including the American Psychological Association, Academy of Management, American Society for Personnel Administrators, and the American Sociological Association. This seemed an appropriate strategy, since management theory crosses all these disciplines. As a result of these involvements, we have broadened the perspective of our research interests and incorporated many ideas and findings from these fields.

Scholl and North feel that we have perpetuated a "major false hypothesis" upon our readers—namely that there are usually differences between student and faculty perceptions of teaching effectiveness. In fact, we sought to test *this hypothesis, and our results do not appear to fly in the face of evidence found in other studies.*

Furthermore, we strongly feel that an investigation into the perceptions of students and faculty regarding the use of student evaluations is relevant, timely and important. Differing perceptions may well have an effect on the validity of the instruments. For example, since students are not typically trained raters, the possibility of error caused by halo effect or contrast effect is very great. If we as teachers are to be rated, does it not seem equitable to have some agreement as to what constitutes good teaching? Must we conclude that disagreement between rater and rated has no effect upon the rating? We think not. Furthermore, an evaluation instrument should include relevant items. Should we merely assume that students know what the relevant attributes to be measured are? Or for that matter, do individuals who construct teacher evaluation instruments systematically consider the issue of relevance? In many cases we think not. These comments are not intended to insult either students or those who construct teacher evaluation instruments. The reality is that many of the instruments are neither reliable nor valid, and most of those who rate others have no training in performing evaluations. (One notable exception to our statement concerning instrument validity and reliability is the Student Instructional Report which has been painstakingly developed and validated by the Educational Testing Service.)

Our conclusions are consistent with the results found in our study. We made every effort not to over-step our bounds in drawing conclusions based upon results from a small, select sample. The "red herring" of federal law enforcement that Scholl and North refer to is our call for an anticipatory, rather than a reactive approach to administrative decision making which includes teacher evaluations as an integral component. Our rationale is simply that the experiences of those in the private sector who have utilized unreliable instruments and/or instruments which were not valid, have been both painful and costly. This is particularly true where prima facie cases of discrimination were presented because the burden of proof shifts to the employer to show that personnel decisions were based on (job related) reliable, valid *instruments (keep in mind that in the EEO area, statistics of underutilization have been sufficient to make a prima facie case for discrimination).*

How many universities can claim that they have validated their teaching evaluation instruments, and that the items contained in the instruments actually measure teaching effectiveness? Can these instruments withstand the close scrutiny of an enforcement agency or, for that matter, an in-house review?

A review of teaching evaluation forms at several universities reveals a wide ranging variety of items differing not only across schools, but across departments within a school. Such differences would not exist if, as Scholl and North suggest, we had settled on a common definition of teaching effectiveness, and if universities were using reliable, valid instruments.

Finally, we do not hold a negative view of the student, nor was there any intention to denigrate students in their role as evaluators of instruction. Labelling students as "subordinates" fit well with the performance evaluation typology employed in our discussion, since we drew our analogy from the business sector. Similarly, we would view administrators as superiors and other faculty members as peers. We do not have a "deplorable perspective" toward students. We both enjoy our work, and have both received college-wide student recognition and awards for outstanding teaching. We both feel that a large part of our enjoyment in teaching is the chance to learn from our students and to grow along with them.

In summary, the research we reported in the Winter, 1979, issue of POD Quarterly *was not a great work; but great works are few and far between. We do feel, however, that it is a careful, methodologically sound study on an important topic, one which deserves further research by us and by others. We were quite frankly shocked by the tone, subjectivity and emotionalism of the Scholl and North letter, and we were surprised by the lack of substantive issues raised. Their comments seem not to be directed toward our research methodology or the validity of the conclusions drawn, but at our failure to cite literature they deem "famous and widely recognized," and the fact that they find our conclusions (and hypotheses) unpalatable.*

We are pleased by the academic environment that allows Scholl and North's letter to be published in a respectable (our view) journal for all the world to see. In this way, the readers can read our study, their remarks and our reply and draw their own conclusions. We are indebted to the editorial staff, Dr. Outcalt, and Ms. Johnson for their assistance, and to the reviewers for their unbiased review of our study.

R. Penny Marquette
George E. Stevens