Differing Student and Faculty Perceptions of Teaching Effectiveness and the Value of Student Evaluations

George E. Stevens
R. Penny Marquette

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

Part of the Higher Education Administration Commons

Stevens, George E. and Marquette, R. Penny, "Differing Student and Faculty Perceptions of Teaching Effectiveness and the Value of Student Evaluations" (1979). POD Quarterly: The Journal of the Professional and Organizational Development Network in Higher Education. 100.
http://digitalcommons.unl.edu/podqtrly/100
Differing Student and Faculty Perceptions of Teaching Effectiveness and the Value of Student Evaluations*

GEORGE E. STEVENS and R. PENNY MARQUETTE

Over the past decade, the use of student evaluations of teacher effectiveness has grown in popularity. The student typically completes a standardized evaluation form for the teachers of classes in which he or she is enrolled. Not only are more schools using this method of assessing teacher effectiveness, many also use the results to make faculty retention, promotion, salary, tenure and other personnel decisions. As Centra (1972) indicates, the question is no longer whether college teaching should be evaluated; it is how, when and by whom? Although various approaches to measuring teaching effectiveness exist (i.e., peer ratings, superior ratings, classroom visits, etc.), the use of student evaluations seems the most controversial. Evidence of the wide-spread use of student evaluations has been documented by a number of researchers (e.g., Peterson, Kerin and Martin, 1978; Lein and Merz, 1978). For example, in a study designed to learn how business faculty were being evaluated, Lein and Merz received responses from 374 business schools. While these schools used widely differing combinations of methods in evaluating faculty, over 70 percent of the schools used some form of student evaluation.

As usage of student evaluations increases, so has the amount of

* The research was supported in part by a grant from The National Fellowships Fund. The authors are grateful for the comments and suggestions of Angelo Kinicki.

POD Quarterly, Vol. 1, No. 4 (Winter 1979)
literature reporting the uses and abuses of these devices (see e.g., Miller, 1978; and Miller and Brokaw, 1978). It is evident that there are both opponents and proponents of the use of student evaluations as input into personnel decisions. Most faculty members are in favor of student evaluations being used for faculty development purposes but are leery of these devices being used for other purposes. One reason for this concern is the many reliability and validity issues related to student evaluations—issues which have been investigated by a number of researchers. Researchers have discovered, for example, that many who construct such ratings are not sufficiently qualified to do so (Costin et al., 1971). Furthermore, when colleague and supervisor ratings of teacher effectiveness were also obtained, low correlations were found between colleague or supervisor ratings and student ratings. Some researchers (e.g., Rodin and Rodin, 1972) find that students are not able to judge teaching effectiveness.

Many variables have been identified which influence student perceptions of teacher effectiveness. In many cases, either the teacher cannot control the variable or the variables may be difficult to measure. Studies undertaken include those examining student attributes such as student achievement (Banziger and Smith, (Note 2); Costin et al., 1971); achievement factors (Banziger and Smith, (Note 2)); personality traits (Warren and O'Connell, 1978); and sex of student (Wilson and Doyle, 1976). Also, a number of other variables have been examined including leader behavior or style (Swanson, 1975; Kinicki and Schriesheim, 1978; Baba and Ace, (Note 1)), type of course, i.e., required vs. elective (Miller, 1978; Miller and Shaaban, 1978), sex of teacher (Elmore and LaPointe, 1975; Wilson and Doyle, 1976), class size (Miller, 1978; Miller and Shaabon, 1978), teacher demands (Sullivan and Skanes, 1974) and teacher personality (Elmore and LaPointe, 1975; Witty, 1947). Although the scope of this paper precludes discussion of these issues, the interested reader is directed to see reviews such as that by Costin, Greenough and Menges, 1971, or Sullivan and Skanes, 1974.

Despite the proliferation of literature on the subject of teacher evaluations, few researchers have studied the reactions of both students and faculty to their usage. One notable exception is the work of Costin, Greenough and Menges, pp. 522–524. Research into student and faculty reactions to the student evaluations procedure would add to our knowledge about the dynamics of this process and fill an important void in our understanding.
The present study is designed to serve this purpose by asking those most affected by the student evaluation process—students and teachers—about their attitudes toward the use of student evaluations and the importance of selected teacher traits. If, as the authors anticipate, significant differences exist among faculty and student ratings, then the potential value of student evaluations becomes suspect. Specifically, ten teacher traits were identified and students as well as faculty members were asked to rate the importance of each trait. The traits selected are the same ones studied by Baum and Brown, (Note 3). Secondly, students and faculty responded to a sixteen item questionnaire concerning the value and use of teacher evaluations. The answers to two broad questions were sought:

1. To what extent do students and teachers agree in their ratings of the importance of a selected set of teacher traits?
2. To what extent do students and teachers agree on the value of student evaluations in assessing teacher effectiveness?

In an attempt to answer these questions, two hypotheses were postulated:

Hypothesis 1. There will be significant differences between faculty members and students on their ratings of the importance of selected teacher traits.

Hypothesis 2. There will be significant differences between faculty members and students in their perceptions of the value and usage of student evaluations of teacher effectiveness.

Method

Sample

The sample consisted of 142 students enrolled in business courses and 55 members of the faculty in the College of Business at Kent State University. The students were chosen at random from six upper division undergraduate courses. All faculty members were asked to participate in the study. The average faculty member respondent is approximately 39 years of age and has attained the rank of assistant professor. This respondent serves on 2–3 committees, published at least one journal article but not a textbook, during his 8½ years of teaching. The average student responding to the survey is about 23 years of age, is a junior with a "B" average, and majors in business administration. This respondent has two years of work experience and currently works approximately 12 hours a week. In-
complete questionnaires were obtained from eight students and five faculty members. These questionnaires were dropped from the analysis, leaving a usable sample of 134 students and 50 faculty.

Procedure

Two questionnaires were administered to the students during regular class time while faculty member questionnaires were placed in campus mail-boxes. The questionnaires included a cover letter assuring the respondent of complete confidentiality. First, the respondent completed a section of the questionnaire which asked for demographic information, e.g., respondent age and sex. Students also gave information about course hours taken, academic major, hours employed per week, etc. Faculty questionnaires asked for additional information such as years of teaching experience, publications to date, academic rank and number of committee assignments during the current year.

Instruments

Teaching Effectiveness. Each respondent completed a short questionnaire listing ten teaching traits that are commonly exhibited in a classroom situation, particularly when a faculty member utilizes a lecture format in presenting material. The instrument was developed by Baum and Brown (Note 3), and employed in their study of student and faculty perceptions of teaching effectiveness (See Table 1). As in the earlier study, each respondent was asked to distribute 100 points across the ten traits according to his/her view of each trait's relative importance in determining teaching effectiveness.

Student Evaluation. Respondent reactions to the use of student evaluations as a measure of teaching effectiveness were assessed using a scale developed by the authors. The instrument consists of 15 statements about what it is that student evaluations measure and how these data should be utilized (See Table 2). Respondents were asked to indicate how strongly they agreed or disagreed with each statement using 7-point Likert scales ("1" = strongly disagree, "7" = strongly agree).

Results

Hypothesis 1 stated that faculty and student responses regarding
the importance of selected teacher traits would be significantly different. Mean scores of both groups were calculated and compared for each of the ten traits. Comparison of these mean with t-tests reveal that the difference between the response means are significant for seven of the ten traits. Six of the seven are significant to the \( p < .01 \) level. Data relevant to Hypothesis 1 (which discusses teacher traits) are presented in Table 1.

<table>
<thead>
<tr>
<th>Teaching Trait</th>
<th>Faculty (N = 50)</th>
<th>Students (N = 134)</th>
<th>t-value</th>
<th>Mean</th>
<th>S.D.</th>
<th>Mean</th>
<th>S.D.</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lectures are easy to outline</td>
<td>8</td>
<td>6.24</td>
<td>4.15</td>
<td>5</td>
<td>11.14</td>
<td>8.61</td>
<td>3.86**</td>
<td></td>
</tr>
<tr>
<td>2. Lectures are entertaining</td>
<td>10</td>
<td>5.90</td>
<td>4.33</td>
<td>6</td>
<td>8.98</td>
<td>6.26</td>
<td>3.20**</td>
<td></td>
</tr>
<tr>
<td>3. Indicates what is important for exams</td>
<td>9</td>
<td>6.02</td>
<td>5.05</td>
<td>3</td>
<td>12.81</td>
<td>6.80</td>
<td>6.43**</td>
<td></td>
</tr>
<tr>
<td>4. Expects students to be prepared</td>
<td>6</td>
<td>8.96</td>
<td>5.41</td>
<td>8</td>
<td>7.46</td>
<td>4.78</td>
<td>1.83*</td>
<td></td>
</tr>
<tr>
<td>5. Emphasizes factual knowledge</td>
<td>5</td>
<td>9.14</td>
<td>6.57</td>
<td>7</td>
<td>8.29</td>
<td>5.27</td>
<td>.91</td>
<td></td>
</tr>
<tr>
<td>6. Emphasizes concepts</td>
<td>1</td>
<td>18.78</td>
<td>9.32</td>
<td>2</td>
<td>13.40</td>
<td>6.31</td>
<td>4.48**</td>
<td></td>
</tr>
<tr>
<td>7. Stresses applications</td>
<td>2</td>
<td>13.86</td>
<td>5.95</td>
<td>1</td>
<td>13.82</td>
<td>7.81</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>8. High grading standards</td>
<td>7</td>
<td>8.86</td>
<td>4.79</td>
<td>10</td>
<td>5.72</td>
<td>4.34</td>
<td>4.24**</td>
<td></td>
</tr>
<tr>
<td>9. Creative thinking on exams</td>
<td>4</td>
<td>9.64</td>
<td>6.23</td>
<td>9</td>
<td>7.01</td>
<td>4.66</td>
<td>3.09**</td>
<td></td>
</tr>
<tr>
<td>10. Exhibits concern for students</td>
<td>3</td>
<td>10.60</td>
<td>4.69</td>
<td>4</td>
<td>11.22</td>
<td>5.54</td>
<td>.70</td>
<td></td>
</tr>
</tbody>
</table>

\( * p < .05 \)

\( ** p < .01 \)

As can be seen, not only are the differences significant, but it is also clear that faculty and students rank the items differently in terms of importance. Specifically, traits 1, 2, and 3 were ranked lowest by faculty but students rated these traits, 5th, 6th and 3rd respectively. Differences also existed in the ranking of other traits; for example, trait 9 was ranked ninth by students but was elevated to 4th place by faculty members. Further examination of these data, however, indicate areas of agreement by faculty and students; for example, traits 6, 7 and 10 are rated similarly by both groups. Additional evidence is provided when the traits are ranked on each group. The Spearman rank correlation between the faculty and stu-
dent ranking is +.419. This positive correlation indicates some general agreement between faculty and students; the correlation, however, is not significant.

Hypothesis 2 stated that faculty and student perceptions of the values and usage of student evaluations would differ significantly. Mean scores of both groups were calculated and compared for each of the 15 items on the student evaluation questionnaire. Comparison of these means by t-tests reveals that the difference between the two groups' responses are significant for eleven of fifteen items. Eight of the differences are significant at the one percent level and three are significant at the five percent level. These results are presented in Table 2.

As can be seen in Table 2, students are in agreement that student evaluations do measure teacher effectiveness. There is far less agreement on the part of faculty, however, that these instruments accomplish this task. The responses on this item differ significantly $t(182) = 3.02, p < .01$. Similarly, students are in strong agreement that they are qualified to evaluate their teachers, while faculty members agree only slightly. The responses on this item also differ significantly $t(182) = 5.70, p < .01$. Both groups disagreed with the statement that evaluations should not be used, but students were significantly stronger in their disagreement $t(182) = 2.60, p < .01$.

One means of determining the extent to which respondents are consistent in their responses is to ask the reverse of a previous question. This tactic was employed in the present study, using the statement, “Only professors have the knowledge to rate their peers.” Both groups disagreed with the statement, although students disagreed more strongly with the group means being significantly different at the five percent level. The companion statement, “Students are qualified to judge professors” was agreed upon by both respondent groups. As noted in the previous paragraph, the responses for the groups were significantly different. This difference results because students had stronger positive feelings about the statement than faculty members. This trend persisted for most of the responses on the questionnaire. Significant differences also resulted on statements related to the value of using student evaluations. Neither group, for example, wanted evaluations discontinued and both groups felt evaluations were an important means of measuring teacher performance. In general, both groups were in agreement or disagreement with the statements provided. The significant differ-
TABLE 2
STUDENT EVALUATION RESPONSES

<table>
<thead>
<tr>
<th>Statement</th>
<th>Faculty (N = 50)</th>
<th>Students (N = 134)</th>
<th>t-value Mean</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Evaluations measure teaching effectiveness</td>
<td>3.44 1.76</td>
<td>4.28 1.65</td>
<td>3.02**</td>
<td></td>
</tr>
<tr>
<td>2. Students use evaluations to punish professors</td>
<td>4.10 1.64</td>
<td>3.75 1.76</td>
<td>1.23</td>
<td></td>
</tr>
<tr>
<td>3. Students are qualified to judge professors</td>
<td>4.10 1.74</td>
<td>5.49 1.36</td>
<td>5.70**</td>
<td></td>
</tr>
<tr>
<td>4. Evaluations serve no useful function</td>
<td>2.22 1.45</td>
<td>2.57 1.67</td>
<td>1.30</td>
<td></td>
</tr>
<tr>
<td>5. Evaluations should not be used</td>
<td>2.94 1.73</td>
<td>2.31 1.36</td>
<td>2.60**</td>
<td></td>
</tr>
<tr>
<td>6. Only professors have the knowledge to rate professors</td>
<td>2.14 1.28</td>
<td>1.81 1.09</td>
<td>1.71*</td>
<td></td>
</tr>
<tr>
<td>7. Student evaluations should be discontinued</td>
<td>2.44 1.72</td>
<td>1.99 1.23</td>
<td>1.98*</td>
<td></td>
</tr>
<tr>
<td>8. Evaluation results should be published</td>
<td>3.40 2.12</td>
<td>4.81 1.86</td>
<td>4.41**</td>
<td></td>
</tr>
<tr>
<td>9. Evaluations are an important means of measuring performance</td>
<td>4.22 1.84</td>
<td>5.27 1.13</td>
<td>4.66**</td>
<td></td>
</tr>
<tr>
<td>10. The more demanding the professor, the lower the rating</td>
<td>3.44 1.86</td>
<td>3.01 1.51</td>
<td>1.59</td>
<td></td>
</tr>
<tr>
<td>11. Evaluations reflect how hard or easy a professor grades</td>
<td>4.26 1.65</td>
<td>3.19 1.49</td>
<td>4.20**</td>
<td></td>
</tr>
<tr>
<td>12. All other performance measures are superior to student evaluations</td>
<td>2.22 1.05</td>
<td>2.71 1.13</td>
<td>2.66**</td>
<td></td>
</tr>
<tr>
<td>13. Student evaluations should not be used for personnel decisions</td>
<td>4.04 1.90</td>
<td>4.72 1.90</td>
<td>2.17*</td>
<td></td>
</tr>
<tr>
<td>14. Evaluations are thrown away and never seen again</td>
<td>3.46 1.88</td>
<td>3.63 1.61</td>
<td>.62</td>
<td></td>
</tr>
<tr>
<td>15. Students and faculty use the same criteria to evaluate performance</td>
<td>1.96 1.19</td>
<td>3.20 1.38</td>
<td>5.62**</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: 7 = strongly agree; 1 = strongly disagree.
* p < .05
** p < .01

ences in the group response means can be attributed to the students' feeling more strongly about many of the items. In only one instance did they feel differently (in terms of direction) about a statement. While students agreed somewhat that students and faculty members used the same criteria to evaluate performance, faculty felt that this was not the case. This difference was significant, \( t (182) = 6.52, p < .01 \).
Discussion

Although it is difficult to generalize from a single study, the results of this research suggest that students and faculty members do differ in the importance assigned to some teacher traits. While the differences found are of statistical significance, it is not so obvious that these differences are of practical significance for the groups under investigation. The calculation of the Spearman rank correlation indicates that there are areas of agreement in each groups’ assessment of teacher traits. One might note, however, that there are traits which are rated in a quite divergent manner. These latter differences suggest that care must be taken in rating teachers on the basis of student evaluations.

An important area of concern is the research evidence available which identifies a number of factors which are not controllable by teachers yet influence the student ratings of teacher performance.

Faculty members’ concern extends beyond the issue of whether these measures are affected by student attributes or other factors beyond their control. Many believe that students give lower ratings to more demanding professors as well as those who are harder graders. Students in the present study did not agree with either of these perceptions, but faculty members did agree with the latter perception (but not the former). These findings might indicate (assuming students are responding honestly) that certain faculty perceptions of how students respond on evaluations are inaccurate. There is agreement by both groups that student evaluations should be used. This would indicate that such ratings have value, if for no other purpose than for providing feedback.

A final important issue of concern to faculty members is how the evaluation results should be used. One surprising finding of the present study was agreement that student evaluations should not be used for personnel decisions. This finding is made even more surprising by the direction of the difference, with students feeling more strongly than faculty that these evaluations should not be used for that purpose, \( t (182) = 2.17, p < .05 \).

Conclusions

Differences do exist in student and faculty perceptions of both important teacher traits and the value and appropriate use of student evaluations. Given these differences, it seems appropriate to exercise great caution in using the results of ratings in making per-
sonnel decisions. The trend is clear; that is, personnel decisions are being made on the basis of such evaluations and that the results of these evaluations are being widely disseminated. In the future, schools must establish the purpose of student evaluations. In the past, they were used more for development purposes and the results were placed in the hands of the teacher. At present, the question of what the ultimate (or perhaps penultimate) objective of faculty instruction actually is still remains. There is apparent confusion and disagreement on this issue. This confusion and disagreement takes on importance when one considers that student evaluations are being used increasingly as criteria for administrative decisions. As indicated by researchers such as Bernardin and Beatty (1979), such use of the evaluations may lead to direct confrontations with various federal enforcement agencies. The reason for this possible confrontation is two-fold: (1) the use of evaluations for administrative decisions may lead to their designation as a selection procedure, (a selection procedure is any measure used as a basis for hiring, promotion, demotion, merit increase or access to training programs), and (2) as a selection procedure, evaluations would fall under the scrutiny of the 1978 Uniform Guidelines on Employee Selection Procedures. It may be wise, therefore, to use student evaluations exclusively for faculty development; otherwise, organizations may be required to demonstrate that adverse impact does not result from the use of such instruments and that said instruments are valid. However, Bernardin and Beatty (1979) cite studies which indicate that validity standards are probably not met with most student ratings. Additionally, these authors cite evidence which indicates the possibility of adverse impact upon minorities and women.

Where might we go from here? Future areas of research would include the use of multiple evaluations: self-ratings, peer ratings, student ratings, and supervisor ratings. Classroom visitations, although objected to by many teachers, should become a part of this process. Recent articles suggest that an emerging issue is that of confidentiality (e.g., Miller, 1978). Until more reliable, valid and useful instruments are developed results should be placed only in the hands of those who need to know and use the information. Finally, from a research perspective, a more systematic approach needs to be taken in identifying the relevant variables which influence ratings and developing as well as testing better instruments. There is a body of industrial and organizational literature that may be useful in
identifying relevant variables. This literature would also give researchers a base from which to work when attempting to measure the variables of interest. Specifically, this literature provides a more diagnostic perspective concerning performance ratings (in the faculty member's case, student evaluations) and the implications such ratings have for other personnel activities and organizational effectiveness. For example, it would appear that industrial organizations, particularly large ones, place greater emphasis upon systematic job analysis, recruitment and placement activities, and the performance appraisal process than do many academic institutions.

Briefly, job analysis is a procedure which allows a manager to identify what each job entails and what kinds of people should be hired for various positions. Many sources (see e.g., Dessler, 1978; Dunnette, 1976; McCormick et. al., 1969) discuss the steps in job analysis, purposes of job analysis and more commonly used job analysis techniques. Job analysis is an important starting point because job analysis information is used for many purposes including personnel activities such as recruitment, selection and placement, compensation, performance, and training. More specifically, this information is the basis upon which recruiters decide who to hire. Job analysis enables a manager to have a clear understanding of what each job entails which then permits him to estimate the value and appropriate compensation for each job. (Compensation is usually tied to the job's required skills, educational level, safety hazards, etc.) One means of determining desired performance in terms of standards to be achieved and activities to be performed is the use of job analysis information. Appraising an employee's performance involves comparing the individual's actual performance with his or her desired performance. The job analysis information may be used to design training and development programs so that actual performance more nearly matches desired performance.

While these latter comments seem tangential to the major thrust of this article, this discussion attempts to provide a linkage between a general discussion of student evaluations and more specific issues related to student evaluations as possible selection devices that impact the administrative process in academia. If, as this author suggests, one perceives student evaluations as selection procedures, then it is appropriate to examine government literature such as the *Dictionary of Occupational Titles* (1965), *Job Analysis: A Guide*
for State and Local Governments (1973), and the Department of Labor's Handbook for Analyzing Jobs (1972). Assuming the reader agrees that student evaluations, when seen as selection devices, do impact a number of personnel related activities, it seems appropriate to suggest organizational literature such as the Handbook of Industrial and Organizational Psychology (1976), personnel texts such as Glueck (1978), Dessler (1978), and Cascio (1978), as well as selected periodicals (e.g., The Personnel Administrator, Personnel, and Human Resources Management). In sum, the major implication of these summary comments is that if we can forget for the moment our organizational context (academia) we may perceive student evaluations as performance ratings rendered by our subordinates and used as input for administrative decisions. Such a view requires that we consider these instruments as an integral part of the selection procedure. As such, there is a large body of literature, both academic and practitioner oriented, that appears both relevant and useful for those whose goal is the proper utilization, development, and appraisal of a key human resource—college and university professors. Finally, the academic institutions which choose to utilize student evaluations as criteria for administrative decisions must consider the legal ramifications of doing so.

REFERENCE NOTES


BIBLIOGRAPHY
