# Student Perspectives on Teaching and its Evaluation

**KARIN J. SPENCER & LIORA PEDHAZUR SCHMELKIN**, *Hofstra University, New York, USA*

ABSTRACT    *The research on student ratings of instruction, while voluminous, has had minimal focus on the perceptions of the students who do the ratings. The current study explored student perspectives on course and teacher ratings as well as some issues related to teaching effectiveness and faculty roles. It was found that students are generally willing to do evaluations and to provide feedback, and have no particular fear of repercussions. However, they have little confidence that faculty or administrators pay attention to the results, and do not even consult the ratings themselves. The students view teaching and advising as the most important roles that should be played by faculty, yet project that faculty, while also viewing teaching as the most important, would rank research above the more student-interactive advising. Canonical correlations among various scales reveal a strong emphasis on such issues of the importance of faculty respect for student views.*

Student ratings of instruction have been in extensive use on college campuses for several decades (e.g. Marsh, 1987; Seldin, 1993), and, despite some inconsistencies and unresolved issues, have been relatively well accepted by researchers and practitioners in the field. The vast research literature has primarily focused on such areas as their reliability, validity, relationships to other variables, and potentially biasing factors, leading to such conclusions as those by Marsh and Bailey (1993), who held that for well-constructed instruments, such ratings are:

> (a) multidimensional; (b) reliable and stable; (c) primarily a function of the instructor who teaches a course rather than of the course that is taught; (d) relatively valid against a variety of indicators of effective teaching; [and] (e) relatively unaffected by a variety of variables hypothesized as potential biases to the ratings. (p. 1)

Among the plethora of studies are many that have uncovered various dimensions used as organising structures in the creation of instruments. Others have explored differences

and still others have then called into question those same differences regarding how students rate instructors based on a great number of particular variables such as age, sex and expected grades. Representative reviews, psychometric presentations and syntheses of the literature include those by Abrami *et al.* (1990); Aleamoni (1981); Cashin (1988, 1995); Centra (1973, 1977, 1993); Coffey & Gibbs (2001); Cohen (1980); Costin *et al.* (1971); Doyle (1983); Feldman (1976, 1989, 1993); Marsh (1987); Overall and Marsh (1980); Patrick and Smart (1998); Theall and Franklin (1990); Wachtel (1998), and Weimer (1990).

Analyses of the psychometric properties are certainly critical for developing confidence in the various instruments, their administration, and the conclusions thereby reached. Yet relatively scant attention has been paid to the broader issues of the purposes and actual uses of the various instruments, and the perspectives of the different parties involved.

There are three main purposes posited for the ratings: a formative one of giving student diagnostic feedback, both positive and negative, to faculty members about their teaching and areas in need of improvement, and two additional purposes of a summative nature: providing administrators with evaluative data used in personnel decisions such as tenure and promotion, and enabling students to reach out to their peers and formalise that blend of information and opinion that circulates in the informal student grapevine.

Empirical research is only now beginning to appear on the three interested parties: faculty, students and administrators. Regarding the students, implicit in all of the literature is the assumption that they answer these anonymous instruments honestly and willingly. However, relatively few studies have attempted to ask the students their general attitudes toward the evaluation—how conscientiously they respond to the questions, how seriously they take the whole process, and what purposes they think are served by the evaluations. Do inquisitive students really make course selections based in part upon the formal judgement of their fellow students? Is their feedback valued, and an agent for change? These are some of the questions explored in the present study.

In those studies that have been conducted, students surveyed have indicated that the evaluations are important and students are qualified to rate their professors, but they are not too optimistic about the overall weight put by administrators and faculty on student opinion (Abbott *et al.*, 1990; Ahmadi, 1981; Ballantyne, 1998; Friedlander, 1978; Marlin, 1987; Miron & Segal, 1986; Tapp, 1985; Wulff *et al.*, 1985).

The few studies that have focused on students' course selection strategies have found that course and teacher ratings information has some impact on students' course selection (e.g., Coleman & McKeachie, 1981; Leventhal *et al.*, 1975). It is important to point out, however, that course selection is more complex and is developmental in nature (Babad *et al.*, 1999). In addition, other aspects, such as whether the course being selected is the first course versus the last course in the selection process, need to be considered. As Babad (2001) concluded: "The applied implication is that feedback consisting of past SRT [student ratings of teaching] data is not sufficient as the informational basis for CS [course selection], and it is but one of the necessary types of information for selecting courses" (p. 489).

The presence of some, but not many, studies that explore the student perspective of evaluations suggests that this is an important issue. Each of the analyses above speaks to some of the issues addressed in the current study, yet each is also slightly different. Some focus on a different culture or very rigid curriculum/professional orientation (Ahmadi, 1981; Ballantyne, 1998; Harnash-Glezer & Meyer, 1991; Marlin, 1987; Miron & Segal, 1986; Neumann & Neumann, 1981), or a community college setting (Tapp,

1985), and all have concentrated separately on either undergraduate or graduate students. The current research makes a point of using a wider lens to include students from several majors and levels of academic experience, in an effort to uncover a more broad-ranging reaction to evaluation.

## Method

### Procedures

A survey was administered by mail to students at a private, suburban comprehensive university in the northeast of the USA enrolling 12 000 students, approximately 7000 of them full-time undergraduates. The administration of the same university-wide Course and Teacher Rating (CTR) instrument is required in all schools, with the exception of the Law School, at the end of each regular semester.

The questionnaire with explanatory cover letter was mailed to a random sample of 500 students, with 125 students from each of the following grade levels: sophomores, juniors, seniors and graduate students. Freshmen were explicitly excluded, since none of them would have had the opportunity to be exposed to the CTR at the time they answered the survey. After an initial mailing and three follow-ups, no statistically significant differences were found due to wave of responses, and an overall 71% ($N = 355$) response rate was achieved. The response rates of individual groups ranged from 80% of the graduate students to 60% of the seniors. Due to missing data or late returns, subsequent analyses were based on 347 usable responses. The final sample, with a few minor exceptions, represented a good cross-section of the student community with regard to sex, year in school, age, commuter or resident status and major.

### Instrument

The instrument consisted of many aspects related to student opinion about the CTRs. Of primary importance are two issues. The first was concerned with the students' reactions to the current Course and Teacher Rating (CTR) instrument at the university. Students were asked to respond on a 7-point agreement/disagreement continuum to various statements tapping their attitudes toward its content and administration (e.g., "I don't write negative comments because I'm afraid the professor will recognize my writing"). A series of statements asked students to indicate the relative importance, on a 5-point scale, of the uses to which the CTR was put and should be put (e.g., "Feedback to teacher about his/her teaching skill"). Also asked were some procedural questions (e.g., who should see the results, if the students had ever read the published summaries of the CTRs).

The second issue concerned students' overall views of the teaching environment at the university. Three sets of questions were used to assess what makes for an effective teacher, teaching in general at the university, and the frequency of certain classroom behaviours. In addition, students were asked about the importance of several roles played by faculty (e.g., teaching, research).

## Results

For each section of the instrument dealing with the analysis of the opinions and attitudes towards teaching and its evaluation, principal axis factor analyses, using squared

TABLE 1. Means and standard deviations on attitude scales

| Scale | Mean | SD |
|---|---|---|
| Reluctance to Do Evaluations[a] | 2.94 | 0.94 |
| Student Opinion Taken Seriously[a] | 4.28 | 1.18 |
| Potential Repercussion Against Students[a] | 2.24 | 1.19 |
| Importance of Indicators of Effective Teaching[b] | 4.36 | 0.50 |
| Impact of Teaching on Students[a] | 4.55 | 0.87 |
| Faculty Openness[c] | 3.17 | 0.61 |

[a] 7-point scale, with 1 = Disagree Very Strongly and 7 = Agree Very Strongly.
[b] 5-point scale, with 1 = Not Important and 5 = Very Important.
[c] 5-point scale, with 1 = Never and 5 = Always.

multiple correlations as initial estimates of communality, were performed. Both the eigenvalues of the reduced matrix and the cumulative proportion of the common factor variance were examined, solutions with more than one factor were rotated both orthogonally and obliquely, and factor correlations from the oblique solution were examined for meaningfulness. For interpretive purposes and to select items for resulting scales, factor loadings $\geq 0.4$ were considered meaningful. Alpha reliability coefficients were estimated on all resultant scales and scoring consisted of averaging the responses for the retained items. Negatively-worded items were recoded.

### Reactions to CTRs and their Use

*Course and Teacher Ratings.* Eighty-four per cent of the respondents indicated that they had filled out the CTRs, and these students responded to 22 items that asked their level of agreement or disagreement with various statements about the CTRs. A 3-factor oblique solution, accounting for approximately 91% of the common factor variance, was retained as the most appropriate.

Seven items loaded on the first factor, which was named Reluctance to Do Evaluations. Reliability for this scale, estimated with Cronbach's alpha ($\alpha$), was 0.67, and scoring produced a mean of 2.94 ($SD = 0.94$; the means on all the attitude scales are summarized in Table 1). This indicated that, as a group, the respondents were not particularly reluctant about doing the evaluations.

Four items loaded on the second factor, which was named Student Opinion Taken Seriously; reliability was estimated at $\alpha = .72$. The mean of 4.28 ($SD = 1.18$) indicated that, as a group, the respondents were ambivalent about their contribution to the evaluation process.

The three items that loaded on the third factor appeared to tap a construct that reflected a concern for potential repercussion, or bias on the part of the instructor, when students submitted their evaluations. The factor was therefore named Potential Repercussion Against Students. Reliability was estimated at $\alpha = 0.70$. The mean of 2.24 ($SD = 1.13$) indicated that, as a group, the respondents did not show much concern about any repercussion against them. It should be noted that the higher standard deviations on the second and third factors suggest a greater divergence of opinion among the respondents.

*Use of Course and Teacher Ratings.* Students were also presented with five purposes for which student ratings of instruction are traditionally used. They were first asked to indicate, on a 5-point scale, their views of how important the CTRs SHOULD BE for

TABLE 2. Means and standard deviations on use of CTRs

| Purpose | Should Be Used | | Really Used | |
|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* |
| Promotion | 3.48 | 1.13 | 2.48 | 1.16 |
| Merit salary increases for teacher | 3.34 | 1.16 | 2.44 | 1.18 |
| Tenure | 3.75 | 1.18 | 2.61 | 1.26 |
| Providing information to other students | 4.10 | 1.05 | 2.71 | 1.38 |
| Feedback to teacher about his/her teaching skill | 4.68 | 0.68 | 3.04 | 1.34 |

*Note*: 5-point scales.

each of those purposes, and then indicate how they think the CTRs are REALLY USED with regard to those same areas. The rank order of the five purposes emerged the same in each question, with feedback to the teacher deemed the most important, followed by the provision of information to other students. However, the means in the SHOULD BE category were always higher than those in the REALLY USED category (see Table 2 for all means and standard deviations). The small standard deviation for feedback in the SHOULD BE question reflects the homogeneity of student opinion in that regard; the relatively low means and high standard deviations for all of the REALLY USED answers reflect a real divergence of opinion on the part of the respondents, and potential pessimism about the overall use of the instrument.

When students were asked their opinions on who should see the results of the evaluations, almost all respondents (96%) chose the Chair, even more than chose the individual professor being evaluated (91%). The third, in their estimation, to see the results were students (79%), followed by the Dean (71%), and in a distant fifth place, the Provost (43%). Despite these findings, students indicated overwhelmingly that they did not consult the published Course and Teacher Ratings, with a majority admitting that they did not even know they were published.

### Views of the Teaching Environment

*Effective Teacher.* Seventeen items asked students to rate the importance of a number of qualities that potentially make for an effective teacher (e.g., "responds clearly to student questions"). A one-factor solution, accounting for approximately 87% of the common factor variance, was retained as the most appropriate, with all 17 items loading meaningfully. This factor was named Importance of Indicators of Effective Teaching. Reliability was estimated at $\alpha = 0.89$. The mean of 4.36 ($SD = 0.50$) indicated that, as a group, respondents felt that these were fairly important indicators of effective teaching.

*Teaching in General.* Fourteen items asked students to indicate their level of agreement or disagreement with a number of statements concerning teaching in general at the University (e.g., "My courses have taught me to apply new principles to solve problems"). A one-factor solution, consisting of seven items and accounting for 73% of the common factor variance, was retained as the most appropriate. This factor was named Impact of Teaching on Students; reliability was estimated at $\alpha = 0.73$. The mean of 4.55 ($SD = 0.87$) indicated that, as a group, the respondents only somewhat agreed that teaching at the university had produced an impact on them.

TABLE 3. Means and standard deviations on faculty roles

|  | Student view | | Projected faculty view | |
| --- | --- | --- | --- | --- |
| Purpose | Mean | *SD* | Mean | *SD* |
| Advising students | 4.34 | 0.85 | 3.46 | 1.20 |
| Research in his/her academic field | 4.03 | 0.94 | 4.16 | 0.95 |
| Service to local community | 2.72 | 1.07 | 2.40 | 1.08 |
| Service to university | 3.60 | 1.02 | 3.40 | 1.14 |
| Teaching | 4.88 | 0.38 | 4.39 | 0.84 |

*Note*: 5-point scales.

*Classroom Behaviour.* Another 14 items asked students to indicate the frequency with which they observed certain types of classroom behaviour, mostly interactions between students and professors (e.g., "My professors bring up viewpoints other than their own"). The one-factor solution, producing eight items with meaningful loadings and accounting for 86% of the common factor variance, was retained as the final solution. After examination of the meaningful items, this factor was named Faculty Openness. Reliability was estimated at $\alpha = 0.77$. The mean of 3.17 ($SD = 0.61$) suggested that, as a group, the respondents felt that only some of the time did their professors exhibit an openness to student opinions and ideas.

*Roles of Faculty—Student View and Projected Faculty View.* Students were presented with a list of five of the more commonly accepted roles that are played by faculty members on a college campus, and asked to indicate on a 5-point scale how they (i.e., STUDENTS) viewed the importance of each of the roles. They were then presented with the same list and asked to project how FACULTY members might rate each of those same roles. Students rated teaching and then advising as the most important, with service to the community as the least important (see Table 3 for means and standard deviations). The small standard deviation on teaching, and the fact that no individual student gave a score of less than 2 to that question, speaks to the overall agreement about its importance. When asked to project about faculty views, however, the rankings changed. Teaching remained the highest, though with a lower mean than in the student view, and research emerged as second in importance. Service to the community was still in last place.

### Tests of Differences

*Attitudes Towards Teaching and Its Evaluation.* The effects of several of the demographic variables on the six attitude variables were explored through MANOVA (multivariate analysis of variance) to determine if there were statistically significant differences based on gender, year, school and full-time/part-time status.

The results of these tests indicate that there were no significant differences in attitudes based on full-time/part-time status ($\Lambda = 0.96302$; $F = 1.78582$; $df = 6, 279$; $p > 0.05$). The MANOVA based on gender was significant ($\Lambda = 0.94844$; $F = 2.59138$; $df = 6, 286$; $p < 0.05$). However, since it only accounted for 5% of the variance, it was deemed not meaningful and therefore not explored in further detail.

Significant and meaningful differences were found due to year ($\Lambda = 0.89042$; $F = 1.87045$; $df = 18, 803.76$; $p < 0.05$), with 11% of the overall variance accounted

for. In order to interpret the nature of these differences, the discriminant function was examined. The dimension reduction analysis indicated that only the first function was significant. Within this function, only Importance of Indicators of Effective Teaching had a meaningful structure coefficient, with 75% of its variability accounted for. Graduate students and seniors represented one cluster of scores, with means of 4.26 ($SD = 0.57$) and 4.28 ($SD = 0.53$) respectively, and the juniors and sophomores were grouped together with slightly higher means of 4.52 ($SD = 0.35$) and 4.42 ($SD = 0.39$) respectively. Although all the means are relatively high, it appears that the more exposure students had to teaching, the less stock they may have put into the importance of these indicators of effective teaching. Sophomores and juniors saw them as more important than did the seniors and graduate students.

Also significant was the school ($\Lambda = 0.90566$; $F = 2.26043$; $df = 12, 534$; $p < 0.05$), with an overall 9% of the variance accounted for. The structure coefficients in the discriminant function indicated that four of the six variables had meaningful loadings. However, upon examination, the most important variable was Faculty Openness, with 94% of its variability accounted for by the discriminant function, followed by Impact of Teaching on Students (31%), Student Opinion Taken Seriously (12%), and Importance of Indicators of Effective Teaching (11%).

In general, the pattern of mean differences revealed the highest means for the students in Education, suggesting that those students, involved in the teaching process themselves, were more optimistic that student opinions are taken seriously and that teaching has an impact on students. Business students produced the lowest means, suggesting their scepticism, with students in the Liberal Arts and Sciences occupying the middle position.

*Use and Role Variables.* The effects of the demographic variables on the use and roles questions were also explored through MANOVA. The results of these tests indicated that there were no significant differences in the student view of the use of CTRs and the roles of faculty members, based on either gender ($\Lambda = 0.90781$; $F = 1.31503$; $df = 20, 259$; $p > 0.05$) or school ($\Lambda = 0.80468$; $F = 1.38884$; $df = 40, 484$; $p > 0.05$).

The MANOVA based on year was significant ($\Lambda = 0.71873$; $F = 1.49755$; $df = 60, 767.58$; $p < 0.05$), with an overall 28% of the variance accounted for. A review of the discriminant function reveals meaningful loadings on three items: Should Be Used—Information to Other Students, Really Used—Information to Other Students, and Roles per Faculty—Research. Means for the first two revealed a similar pattern, with sophomores being the most idealistic about the provision of information to other students; scores for juniors, seniors, and graduate students suggested that the longer students were on campus, the more sceptical they became. With respect to the research role for faculty, graduate students projected that faculty would rate research as a fairly important role, whereas undergraduates produced lower means.

The MANOVA based on part-time/full-time status was also significant ($\Lambda = 0.87097$; $F = 1.88151$; $df = 20, 254$; $p < 0.05$), with an overall 13% of the variance accounted for. The meaningful loadings on the discriminant function were for the real and potential use of the CTRs as sources of information to students, the student view of how faculty perceive their roles of service to the community and to the institution, and the use of the CTRs in the tenure process. In all instances, full-time students displayed higher means than their part-time counterparts, suggesting that these issues were more important to them.

*Relationships Among Variables*

An examination of the correlations among the attitude variables revealed that Impact of Teaching on Students correlated with Faculty Openness ($r = 0.52$), suggesting that the more positively students felt about teaching at the university, the more frequently they felt that faculty were open to student concerns and interacted with students. An additional correlation of Impact of Teaching with Student Opinion Taken Seriously ($r = 0.36$) further suggested that the more positively they felt about teaching, the more they believed that their opinions counted. The two negative factors that came from the CTR questions, Reluctance to Do Evaluations and Potential Repercussion Against Students correlated with one another ($r = 0.37$), suggesting that students concerned about repercussions were also more reticent to do the evaluations.

In order to examine the relationship between the six attitude variables (Set A) and the 20 Use and Roles questions (Set B), a canonical correlation analysis was performed. The results of this analysis yielded statistical significance ($\Lambda = 0.30021$; $F = 2.80123$; $df = 120$, $1452.31$; $p < 0.05$). The redundancy analysis suggested that only the results associated with first canonical correlation should be examined with 14% of the variation in Set A predictable from Set B.

For this first canonical correlation, the linear combinations of the two sets of variables share 45% of the variability. An examination of the structure coefficients of Set A revealed high loadings on Impact of Teaching on Students, Faculty Openness, and Student Opinion Taken Seriously, all scales that clearly share the concept of respect for students' opinions and the importance of interaction with students. Set B yielded meaningful correlations on items dealing primarily with faculty–student interaction, focusing on the non-research roles of professors and the use of CTRs as feedback to both students and professors. Therefore, overall student opinions and attitudes seem to be correlated mostly with items that stress interaction.

## Discussion

Students were ambivalent about whether or not their opinions expressed on the CTRs were taken seriously, yet were not very reluctant to evaluate their faculty members, and overall, showed little fear of repercussion against them in the administration of the ratings forms. Students also agreed that various indicators of effective teaching were all very important. They somewhat agreed that teaching at the university had made a positive impact on them, but when presented with a series of descriptions of faculty–student interactions and examples of faculty openness to student views, they were fairly non-committal about how frequently these behaviours occurred.

The attitudes towards the CTRs divided themselves into three distinct factors, two of which reflected potential negative aspects of doing the evaluations. These were Reluctance to Do Evaluations and Potential Repercussion Against Students. However, the means on both of these scales indicate that students are not particularly reluctant to do the evaluations, nor are they overly concerned about the potential repercussion against them. The latter finding is surprising, given the literature that stresses students' desire for anonymity (e.g., Braskamp & Ory, 1994; Feldman, 1979).

The third scale, Student Opinion Taken Seriously, reflected students' ambivalence as to whether or not their opinions are taken seriously. These findings of uncertainty as to the importance of student opinion are also substantiated by the optional comments made by students at the end of the survey, as well as the results of the relatively few studies

that have addressed this issue (Ballantyne, 1998; Miron & Segal, 1986; Tapp, 1985; Wulff *et al.*, 1985).

The items that loaded on the various factors of teaching effectiveness and faculty interaction all reflected the major issues brought out in earlier research. In the Importance of Indicators of Effective Teaching scale, issues of clarity, fairness and respect for students were paramount, items also highlighted in Impact of Teaching on Students. The issue of faculty concern for students confirms Feldman's 1976 findings that among the highest rated of his dimensions were those of instructor's concern or respect for students. Like the Impact of Teaching on Students scale, items that loaded high on Faculty Openness are corroborated by research emphasising the contribution of interactions and open communication towards a general satisfaction with one's education (Harnash-Glezer & Meyer, 1991; Light, 1990, 1992).

Overall, in the current study, the items with the highest loadings emphasised students' perceptions of a willingness on the part of faculty to interact with them, in accommodating their special needs, giving feedback, sensing when they were having trouble with the material, and knowing them by name. The items also reflected an openness and respect by faculty for students as fellow learners, in bringing up other viewpoints, admitting when they do not know something, giving value to student opinion, and seeking feedback from students in matters regarding their teaching style.

The exploration of the potential effect of various background or other variables on the six attitude scales and additional major questions primarily showed differences in two different groupings of students: (a) the school with which a student's major department is affiliated and (b) year in school. These results are worthy of note.

Marsh cautions against putting too much store on background variables as biasing, commenting that "perhaps more than any other area of student evaluation research, the search for potential biases is extensive, confused, contradictory, misinterpreted, and methodologically flawed" (Marsh, 1983, p. 151). This is especially dangerous in the absence of an explicit definition of bias (Marsh, 1984). However, in the current study, the issue is not one of background variables contributing to bias in how individual students respond to actual questions in the CTRs. Rather, it focuses on potential differences that affect how students approach or perceive the evaluations, and reflects an overall recognition of (a) how a student's affiliation with a major affects his/her general perception of teaching and classes—be they classes in the major department or not; and (b) how the length of time the student has been part of the educational process influences his/her views of these same variables. In addition, the differences found, while meaningful and significant, were not so overwhelmingly large as to raise extreme concerns about their detrimental effects on the evaluation process.

The differential by school reflected the fact that students in Business scored considerably lower than their counterparts in Arts and Sciences and Education. All three schools have substantially different environments and provide students with different choice patterns in course selection. Business students typically have a more structured program with a higher percentage of larger classes, a factor that can limit the amount of interaction a faculty member is able to have with his/her students. Year in school differences revealed a pattern of growing scepticism about how seriously student opinion is considered; the longer a student has been on campus, the less idealistic he or she will be regarding the use of the CTRs.

The intercorrelations of the various scales and the Use of Course and Teacher Ratings questions confirmed the clusters of student priorities and opinions. Those students who were concerned about repercussions were also those who were more reticent about doing

evaluations. On the other hand, the more positively students felt that teaching had affected their lives, the more frequently they felt that faculty were interactive and open to diverse viewpoints, and the more those students thought that their opinions were taken seriously. These were also the students who scored higher on questions that emphasised the importance of feedback from students as a major use of student ratings.

The results of the overall survey confirm that students are sceptical about the use of the ratings as a barometer of student opinion about professors and classes. They are not reluctant to do them, nor do they have any particular fear of bias. However, since they are unsure whether their opinions matter, or to what purpose the ratings are put, they may not pay much attention to them. This is in stark contrast to their demonstrated desire to express their opinions and have an influence on teaching. Their wish to have an impact but their lack of (a) confidence in the use of the results; and (b) knowledge of just how to influence teaching, is reflected in the observation that they do not even consult the public results of student ratings.

The problem therefore seems to be one of convincing students that their opinions do matter. Such a process must be done on several levels. On a surface level, general publicity about student ratings, their purpose and their availability for student review would certainly help. Faculty members must also be apprised of the ambivalence of students toward the ratings, and the general cynicism about the importance of student input.

The larger question, however, is if a computerised end-of-semester instrument should be the only formalised way students have to express their views. In order for students to really feel that connection with their comments and feedback to faculty, there must be some tangible immediacy to the results. Optional student comments in this survey spoke to the recognition that in many cases, students will not have that professor again, and therefore expressing their opinions does not, in the long run, benefit that particular student anyway.

The literature speaks compellingly of the benefits of mid-term formative evaluations where students have an opportunity to express their views and to see possible changes during the remainder of the semester (Abbott *et al.*, 1990; Friedlander, 1978; Overall & Marsh, 1979; Wulff *et al.*, 1985). The end-of-semester evaluations would still be needed for more summative purposes of reappointment, promotion and tenure, as well as feedback to other students. These later evaluations would give students the opportunity to give an overall assessment of the class, something they might not have enough perspective for at the middle of the term.

Instructors can make a point of carrying over feedback from one semester into another, by announcing at the beginning of a course that they are trying a new approach, based upon comments of their students in earlier classes. This feedback could come from both the mid-semester and end-of-semester evaluations, and it is critical that the faculty member mention the source as further reinforcement of the importance of student opinion.

Further research needs to be conducted into the other two constituencies: the faculty perspectives and administrative perspectives. There is some research on faculty attitudes, albeit mostly anecdotal, that suggests that many faculty members are distrustful of student ratings, perceiving students not to be qualified to sit in judgement, and expressing concern that ratings are mere popularity contests (Aleamoni, 1987; Costin *et al.*, 1971; Marsh & Ware, 1982; Ware & Williams, 1975). However, Schmelkin *et al.* (1997) did not find a great deal of resistance on the part of faculty to student ratings in general, nor to their use for formative and summative evaluation. As for administrators—

those individuals in a position to use the summative evaluations for decisions on rank and tenure—an exploration of their actual use would round out the picture of the overall use of these ubiquitous instruments.

## Notes on Contributors

Dr KARIN J. SPENCER is Dean of University Advisement at Hofstra University. *Correspondence:* Dr Karin J. Spencer, Advisement Office, Hofstra University, Hempstead, NY 11549, USA. Tel: (516) 463–6776. Fax: (516) 564–6674. E-mail: advkjs@hofstra.edu.

Dr LIORA PEDHAZUR SCHMELKIN is Professor of Psychology and the Leo A. Guthart Distinguished Professor of Teaching Excellence at Hofstra University. She also directs the University's Center for Teaching and Scholarly Excellence.

## REFERENCES

ABBOTT, R. D., WULFF, D. H., NYQUIST, J. D., ROPP, V. A. & HESS, C. W. (1990) Satisfaction with processes of collecting student opinions about instruction: The student perspective, *Journal of Educational Psychology*, 82 (2), pp. 201–206.

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 (2), pp. 219–231.

AHMADI, G. (1981) A Study of Student Opinions Toward Faculty Evaluation by Students. Unpublished doctoral dissertation, Kansas State University.

ALEAMONI, L. M. (1981) Student ratings of instruction, in: MILLMAN, J. (Ed.) *Handbook of Teacher Evaluation* (Beverly Hills, CA, Sage).

ALEAMONI, L. M. (1987) Typical faculty concerns about student evaluation of teaching, in: ALEAMONI, L. M. (Ed.) *Techniques for Evaluating and Improving Instruction* (New Directions for Teaching and Learning, No. 31, San Francisco, CA, Jossey-Bass).

BABAD, E. (2001) Students' course selection: Differential considerations for first and last course, *Research in Higher Education*, 42 (4), pp. 469–492.

BABAD, E., DARLEY, J. M. & KAPLOWITZ, H. (1999) Developmental aspects in students' course selection, *Journal of Educational Psychology*, 91 (1), pp. 157–168.

BRASKAMP, L. A. & ORY, J. C. (1994) *Assessing Faculty Work: Enhancing Individual and Institutional Performance* (San Francisco, CA, Jossey-Bass).

BALLANTYNE, C. (1998) *Improving university teaching: Responding to feedback from students.* Paper presented at the Adult Learning Cultures: Challenges and Choice Conference, Wellington Polytechnic, New Zealand. Retrieved November 22, 2001, from *http://cleo.murdoch.edu.au/evaluation/pubs/confs/respfeed.html*

CASHIN, W. E. (1988) *Student Ratings of Teaching: A Summary of the Research* (IDEA Paper No. 20, Manhattan, Kansas, Kansas State University, Center for Faculty Evaluation and Development).

CASHIN, W. E. (1995) *Student Ratings of Teaching: The Research Revisited* (IDEA Paper No. 32, Manhattan, Kansas, Kansas State University, Center for Faculty Evaluation and Development).

CENTRA, J. A. (1973) Effectiveness of student feedback in modifying college instruction, *Journal of Educational Psychology*, 65 (3), pp. 395–401.

CENTRA, J. A. (1977) Student ratings of instruction and their relationship to student learning, *American Educational Research Journal*, 14 (1), pp. 17–24.

CENTRA, J. A. (1993) *Reflective Faculty Evaluation* (San Francisco, CA, Jossey-Bass).

COFFEY, M. & GIBBS, G. (2001) The evaluation of the Student Evaluation of Educational Quality Questionnaire (SEEQ) in UK higher education, *Assessment & Evaluation in Higher Education*, 26 (1), pp. 89–93.

COHEN, P. A. (1980) Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings, *Research in Higher Education*, 13 (4), pp. 321–341.

COLEMAN, J. & MCKEACHIE, W. J. (1981) Effects of instructor/course evaluations on student course selection, *Journal of Educational Psychology*, 73 (2), pp. 224–226.

COSTIN, F., GREENOUGH, W. T. & MENGES, R. J. (1971) Student ratings of college teaching: Reliability, validity, and usefulness, *Review of Educational Research*, 41 (5), pp. 511–535.

DOYLE, K. O., JR. (1983) *Evaluating Teaching* (Lexington, MA, Lexington Books).

FELDMAN, K. A. (1976) The superior college teacher from the students' view, *Research in Higher Education*, **5** (3), pp. 243–288.

FELDMAN, K. A. (1979) The significance of circumstances for college students' ratings of their teachers and courses, *Research in Higher Education*, 10 (2), pp. 149–172.

FELDMAN, K. A. (1989) The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies, *Research in Higher Education*, 30 (6), pp. 583–645.

FELDMAN, K. A. (1993) College students' views of male and female college teachers: Part II—evidence from students' evaluations of their classroom teachers, *Research in Higher Education*, 34 (2), pp. 151–211.

FRIEDLANDER, J. (1978) Student perceptions on the effectiveness of midterm feedback to modify college instruction, *Journal of Educational Research*, 71 (3), pp. 140–143.

HARNASH-GLEZER, M. & MEYER, J. (1991) Dimensions of satisfaction with collegiate education, *Assessment and Evaluation in Higher Education*, 16 (2), pp. 95–107.

LEVENTHAL, L., ABRAMI, P. C., PERRY, R. P. & BREEN, L. J. (1975) Section selection in multi-section courses: Implications for the validation and use of teacher rating forms, *Educational and Psychological Measurement*, 35 (4), pp. 885–895.

LIGHT, R. L. (1990) Explorations with students and faculty about teaching, learning, and student life, *The Harvard Assessment Seminars: First report—1990* (Cambridge, MA, Harvard University Graduate School of Education).

LIGHT, R. L. (1992) Explorations with students and faculty about teaching, learning, and student life, *The Harvard Assessment Seminars: Second report—1992* (Cambridge, MA, Harvard University Graduate School of Education and Kennedy School of Government).

MARLIN, J. W., JR. (1987) Student perception of end-of-course evaluations, *Journal of Higher Education*, 58 (6), pp. 704–716.

MARSH, H. W. (1983) Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics, *Journal of Educational Psychology*, 75 (1), pp. 150–166.

MARSH, H. W. (1984) Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility, *Journal of Educational Psychology*, 76 (5), pp. 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 (3), pp. 253–388.

MARSH, H. W. & BAILEY, M. (1993) Multidimensional students' evaluations of teaching effectiveness: A profile analysis, *Journal of Higher Education*, 64 (1), pp. 1–18.

MARSH, H. W. & WARE, J. E., JR. (1982) Effects of expressiveness, content coverage, and incentive on multidimensional student rating scales: New Interpretations of the Dr. Fox effect, *Journal of Educational Psychology*, 74 (1), pp. 126–134.

MIRON, M. & SEGAL, E. (1986) Student opinion on the value of student evaluations, *Higher Education*, 15 (3–4), pp. 259–265.

NEUMANN, Y. & NEUMANN, L. (1981) Determinants of students' satisfaction with course work: An international comparison between two universities, *Research in Higher Education*, 14 (4), pp. 321–333.

OVERALL, J. U. & MARSH, H. W. (1979) Midterm feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes, *Journal of Educational Psychology*, 71 (6), pp. 856–865.

OVERALL, J. U. & MARSH, H. W. (1980) Students' evaluations of instruction: A longitudinal study of their stability, *Journal of Educational Psychology*, 72 (3), pp. 321–325.

PATRICK, J. & SMART, R. M. (1998) An empirical evaluation of teacher effectiveness: The emergence of three critical factors, *Assessment & Evaluation in Higher Education*, 23 (2), pp. 165–178.

SCHMELKIN, L. P., SPENCER, K. J. & GELLMAN, E. S. (1997) Faculty perspectives on course and teacher evaluations, *Research in Higher Education*, 38 (5), pp. 575–592.

SELDIN, P. (1993) The use and abuse of student ratings of professors, *The Chronicle of Higher Education*, 21 July, p. A40.

TAPP, W. M. (1985) An Analysis of Student Opinions Regarding Faculty Evaluations in Selected Community Colleges. Unpublished doctoral dissertation, University of Houston.

THEALL, M. & FRANKLIN, J. (1990) Student ratings in the context of complex evaluation systems, in: THEALL, M. & FRANKLIN, J. (Eds) *Student Ratings of Instruction: Issues for Improving Practice* (New Directions for Teaching and Learning, No. 43, San Francisco, CA, Jossey-Bass).

WACHTEL, H. K. (1998) Student evaluation of college teaching effectiveness: A brief review, *Assessment & Evaluation in Higher Education*, 23 (2), pp. 191–211.

WARE, J. E., JR. & WILLIAMS, R. G. (1975) The Dr. Fox effect: A study of lecturer effectiveness and ratings of instruction, *Journal of Medical Education*, 50 (2), pp. 149–156.

WEIMER, M. (1990) *Improving College Teaching: Strategies for Developing Instructional Effectiveness* (San Francisco, CA, Jossey-Bass).

WULFF, D. H., STATON-SPICER, A. Q., HESS, C. W. & NYQUIST, J. D. (1985) The student perspective on evaluating teaching effectiveness, *ACA Bulletin*, 53 (August), pp. 39–47.