

Student evaluation of instruction: What can be learned from students' written comments?

Fadia Nasser-Abu Alhija^{a,*}, Barbara Fresko^b

^a School of Education, Tel Aviv University, Tel Aviv, Israel

^b Beit Berl Academic College, Israel

ABSTRACT

Extensive research has been done on student ratings of instruction on closed-ended questionnaires, but little research has examined students' written responses to open-ended questions. This study investigated the written comments of students in 198 classes, focusing on their frequency, content, direction, and consistency with quantitative ratings on closed-ended items. Results indicated that about 45% of the students wrote comments. Comments were more often positive than negative and tended to be general rather than specific. Written comments addressed dimensions similar to those identified in the closed-ended items, but they also related to unique aspects of the courses as well.

© 2009 Elsevier Ltd. All rights reserved.

Faculty evaluation is one of the most common evaluation activities routinely carried out in higher education institutions. Its main purpose is two-fold: to motivate instructional improvement and to assist decision-making regarding tenure and advancement of individual faculty members. Although various procedures have evolved for carrying out faculty evaluation (such as student evaluation, peer evaluation, self-evaluation, and administrative evaluation, see McGee, 1995), there seems to be an overly abundant use of student ratings when the focus of evaluation is instruction. Intensive use of student ratings for evaluating instruction in higher education has led to extensive research of both their psychometric qualities (reliability and validity) and their relationship to potential biasing factors (e.g., Obenchain, Abernathy, & Wiest, 2005; Renaud & Murray, 2005; Watchel, 1998). Studies have focused on questionnaires made up of a series of Likert-type items which address different aspects of instruction. Although this type of closed-ended questionnaire is the most common, there exist two other kinds of course evaluation questionnaires: (1) questionnaires on which students write freely in response to open-ended questions and (2) questionnaires which contain a combination of both closed-ended and open-ended items (Sheehan & DuPrey, 1999). Both types contain items that generate data in the form of written comments.

Despite the extensive research literature on student ratings, little is known about the quality of data obtained from students' written comments, their content, and the relationship between them and other variables. The few studies which have been reported in the literature tend to focus on the frequency of

comments, their length, direction, and content, and the characteristics of the students who write them. When relating to the frequency of written student comments great variation has been noted. Some studies have reported that only 10–12% of the students wrote comments (Theall & Franklin, 1991), in other studies the proportion of students writing comments was 40–50% (Hardy, 2003; Zimmaro et al., 2006), while in still others this rate reached as high as 60–70% (Oliver, Tucker, & Pegden, 2007).

Both organizational and individual variables have been noted to have an effect on the rate of commenting. For example when course evaluation is computerized and done online, a tendency has been noted for students to write more comments and longer comments than when evaluation is carried out using paper and pencil (Sorenson & Reiner, 2003). Moreover when a short form of course evaluation is used students were found to write more comments (Johnson, 2003). With regard to individual student characteristics associated with writing comments, Oliver et al. (2007) reported that high achievers, females, older students, native (as opposed to foreign) students, and full-time students wrote more comments than other cross-sections of the student population.

On the whole students tend to write more positive comments than negative comments (Braskamp, Ory, & Pieper, 1981 in Centra, 1993; Hardy, 2003; Oliver et al., 2007; Zimmaro et al., 2006). In fact, in nearly all studies which addressed this issue, positive comments were twice as common as negative comments. Zimmaro and colleagues (2006) reported that positive comments tended to be more general in nature, while negative comments were more specific and focused on specific aspects of a course.

It has been noted that the content of students' comments reflect the dimensions of instruction often identified from factor analyses of results from closed-ended instruments (e.g., Braskamp, Ory, &

* Corresponding author.

E-mail address: fadia@post.tau.ac.il (F.N.-A. Alhija).

Pieper, 1981 in Centra, 1993; Ory, Braskamp, & Pieper, 1980). Teaching competences and interpersonal skills of the instructor are frequently mentioned in students' comments as well as course content, organization, and general quality (Braskamp, Ory, & Pieper, 1981 in Centra, 1993; Oliver et al., 2007; Ory et al., 1980; Zimmario et al., 2006). The literature emphasizes the similarity in the domains addressed by both written comments and student ratings and tends to ignore the differences between them. The question that should be asked is what important additional information can be gained from using questionnaires in which both open-ended and closed-ended items are included?

Schmelkin-Pedhazur, Spencer, and Gellman (1997) found that, in general, faculty members view course evaluation by students as useful. It is likely that the utility of students' written comments is not identical to the utility of the numerical ratings that are generated from their responses to close-ended questionnaire items. The literature seems to suggest that written comments have greater potential for influencing instructors. For example, some studies have indicated that instructors tend to prefer receiving written comments from students as opposed to statistical summaries (Centra, 1993; Lewis, 2001; Svinicki, 2001), and claims have been made that comments from students are more informative (Nasser & Fresko, 2002; Smith & Welicker-Pollack, 2008), more specific, and often contain concrete suggestions for improvement (Hammond, Taylor, & McMenamin, 2003).

As pointed out by Symons (2006), repeated patterns in students' comments can provide useful insight into the issues that are important to them. However, interpreting students' comments is not always as easy as it may seem. Consequently articles have been published suggesting various ways in which instructors can organize and analyze students' comments in order to better learn from them so as to improve instruction (e.g., Hodges & Stanton, 2007; Lewis, 2001; Svinicki, 2001).

The purpose of the present study is to add to this relatively small body of knowledge surrounding open-ended questions on course evaluation questionnaires. Since most of the knowledge that we have about students' written comments derive from studies conducted in either the United States or Australia, it is of interest to know whether students in other countries have similar patterns of response. At the Israeli college where this study took place, the course evaluation instrument was of the mixed type, containing both Likert-type items and open-ended questions. Students' responses to the open-ended questions were analyzed in order to answer the following questions:

1. To what extent do students respond to the open-ended questions and provide comments about the courses?
2. What are the areas and dimensions of instruction most frequently commented upon?
3. What is the direction of student comments (positive or negative)?
4. To what extent are student comments consistent with numerical ratings given on the closed section of the questionnaire and to what extent do they provide additional information not addressed by questionnaire items?

Method

Sample

Comments written by students in 198 different undergraduate courses in an Israeli teacher training college that grants a 4-year bachelor degree in education were analyzed. One course per instructor was randomly selected, which at the time of the study was the policy of the college with respect to course evaluation. In other words, the data collected in this study were part of the

standard evaluation conducted at this college. On the average, 70% of students per course completed the course evaluation questionnaire. The exact number of students who participated in the study is unknown since some students were undoubtedly studying in more than one of the courses. Overall, 3067 questionnaires were analyzed.

Courses in the sample include both introductory and advanced courses in the educational sciences, social sciences, natural sciences, and humanities, as well as professional training courses. About half of the courses were of one semester duration, while the others were yearlong. In general class size in the college tends to be small and the size of the classes in the research sample ranged from 4 to 64 ($M = 15.6$, $SD = 7.23$). The student population at this college is predominately female and in 30% of the courses in the sample all students were female.

The instructors in these courses ($N = 198$) were two-thirds female, with 23 years of teaching experience on the average. Approximately 78% were tenured faculty and half held a doctoral degree.

Procedure

The course evaluation questionnaire was administered during class time by administrative staff during the last 2 weeks of each semester during one academic year. Students' anonymity was maintained: they were not required to identify themselves on the questionnaire form and, in accordance with college policy, their written comments were typed so that they could not be identified by their handwriting.

Research instrument

The course evaluation questionnaire was made up of two parts. In the first part students were presented with 17 Likert type items: 16 items related to various dimensions of instruction and one item was general. They were asked to indicate to what extent each statement described what happened in the course on a 7-point scale from 1 (*very little*) to 7 (*very much*). The second part of the questionnaire included two open-ended questions, which asked students to add any comments on different aspects of the course and to give specific recommendations for improving the course. An entire page was provided for their responses.

Research variables

Written comments

Preliminary screening of students' responses to the two open-ended questions indicated that they did not differentiate between them, making both comments and recommendations in both allotted spaces. Therefore comments and recommendations were combined and analyzed as one data set. What each student wrote was decomposed into meaning units, which were sorted into categories and coded according to direction (positive or negative) and intensity (from +2 to −2). This procedure was carried out by three independent raters in four stages.

Stage 1: A coding scheme was developed based on a preliminary analysis of one-third of the comments. Each rater independently decomposed the comments into meaning units. Results were pooled and 45 content categories were identified, which related to the focus (course, instructor, or context), the primary content area (e.g., teaching style, instructor personal attributes, and course content), and the secondary content area (e.g., interest generated, structure, clarity, and difficulty). A description of the 45 categories and examples of meaning units are presented in Appendix A.

Each meaning unit was assigned a score ranging from −2 to +2. A meaning unit which expressed a *very positive* evaluation was

assigned a score of +2; a *mildly positive* meaning unit received a score of +1; a *mildly negative* meaning unit was assigned a score of –1; and a *strongly negative* meaning unit was assigned a score of –2. Extreme scores (± 2) were assigned in three different situations: when the meaning unit contained strong descriptors (fantastic and horrible); when the meaning unit included superlatives (“the best”, “very”); and when several meaning units referred to the same category (and then that category was coded only once with an extreme score).

Stage 2: Using the above 45 categories, the three raters working together decomposed and analyzed students’ comments on 12 courses. The purpose was to increase uniform application of the coding scheme.

Stage 3: Inter-rater reliability was examined for an additional 25 courses which were independently rated by each rater using the coding scheme. Inter-rater reliability was calculated for three parameters: classification of meaning units into categories, the intensity of the score assigned to them, and its direction. The percentage of agreement among at least two of the three raters was

92% for classification into categories, 98% for intensity, and 100% for direction. These results were considered satisfactory.

Stage 4: Content analysis and coding of the remaining comments were completed for all courses in the sample.

For each of the 45 categories an aggregate score was computed for each course by summing the scores for all the students on that category and dividing by the number of students in the course who had written comments of any kind. In the case in which no student related to a particular category, the score assigned to the course in this category was zero. In this manner, 45 aggregate variables were created.

The 45 categories were conceptually grouped into eight dimensions: three related to the course (general evaluation of the course, course content, course assignments), three related to the instructor (general evaluation of the instructor, personal traits, teaching style) and two related to the context (scheduling issues, student composition). A score was created for each dimension by averaging the specific category scores. In addition, a general score was calculated by averaging the scores on the six most frequently

Table 1

Mean percentage of students in course who commented, percentage of courses in which students commented, and mean coded score per category.

Focus	Primary content area	Secondary content area	Mean percentage of students who commented on category	Percentage of courses in which students commented on category	Mean coded score on the category
Instructor	General Teaching	General evaluation	23.12	67.7	+.387
		Interest	14.87	49.0	+.113
		Clarity	10.04	41.4	+.070
		Structure	8.66	35.9	–.014
		Classroom environment	8.26	37.4	+.078
		Diversity of methods	6.83	31.8	–.011
		Contribution to learning	5.23	26.3	+.068
		Activating students	5.20	23.7	+.029
		Efficient time use	4.70	20.7	–.037
		Combining theory and practice	4.44	19.7	–.003
		Use of teaching aids	3.49	16.2	+.009
		Pace	1.59	8.6	–.019
		Supervision quality	1.01	6.1	–.001
	Personal traits	Conduct towards students	8.94	38.4	+.052
		Flexibility and consideration	8.31	37.4	+.032
		Professional competence	6.72	32.8	+.088
		Commitment	4.06	19.2	+.044
		Emotional support	1.90	10.6	+.018
		Ability to influence students	1.83	8.1	+.022
		Responsiveness	1.63	8.6	+.021
		Sense of humor	.71	4.0	+.012
Course	General Content	General evaluation	15.23	51.5	+.242
		Interest	16.87	56.1	+.178
		Contribution to learning	16.29	53.5	+.111
		Scope	7.74	34.3	–.028
		Relevance	6.96	31.8	–.030
		Compulsory	5.49	27.3	–.048
		Importance	5.40	28.8	+.064
		Difficulty	3.28	15.7	–.039
		Match with syllabus	1.01	4.0	–.011
		Combining theory and practice	1.01	5.1	–.007
	Assignments	Quantity	3.18	15.2	–.039
		Relevance	2.95	15.7	–.026
		Difficulty	2.67	15.7	–.041
		Clarity	2.63	14.1	–.024
		Contribution to learning	1.35	8.6	+.014
		Timing	1.17	5.1	–.004
Context	Scheduling issues	Length of course	3.12	10.1	–.029
		Hour of class meetings	.93	3.5	–.010
		Sequence in training program	.51	.5	–.005
	Student composition	Class size	1.40	6.6	–.020
		Heterogeneity	.57	4.5	–.006
		Discipline problems	.57	4.5	–.008
		Overall academic level	.29	3.0	–.001
		Overall motivation	.10	.5	–.001

mentioned dimensions. Not included in this general score were the contextual categories which were very rarely commented upon.

Student ratings

Principle axis factoring with oblique rotation was conducted on responses to the 16 closed items, resulting in three factors: course content, instruction, and teacher–student relations. Results of the factor analysis are presented in [Appendix B](#). Structure validity pertaining to the rating scale was established in two previous studies (Nasser & Fresko, 2002; Nasser & Hagtvet, 2006). In all cases the same factor structure was obtained.

For each factor an index score was created by averaging the ratings on the items that loaded primarily on that factor. Cronbach reliability coefficients were .95, .83 and .91, respectively. In addition, two other scores were included in the analysis: the first was the score on a single item which measured overall evaluation of the course (*This course was very good*), and the second was an overall mean of the responses on the 16 specific items. This last score had an overall Cronbach reliability coefficient of .95.

Findings

Frequency of written comments

On the average 44.8% of the students per course wrote comments ($SD = 21.1\%$). The percentage of students writing comments ranged from 5% to 100%. Moreover, the percentage of students writing comments was found to be correlated to student ratings on the closed section of the questionnaire. Students in highest and lowest rated courses tended to write more comments. When courses were sorted into quartiles according to student ratings, students in courses belonging to the highest and lowest quartiles wrote more comments (53% and 46%, respectively) than those in the second and third quartiles (38% and 42%, respectively). These differences were statistically significant ($F = 4.28, p = .006$).

What do students write?

Three measures were calculated for each of the 45 categories:

1. The mean percentage of students in each course who wrote comments.
2. The percentage of courses in which the category was mentioned.
3. The mean category score from which the magnitude and direction of the comments could be ascertained.

[Table 1](#) summarizes the results. Findings indicate that only six of the 45 categories were mentioned by at least 10% of the students

on the average. These categories were *general evaluation of the instructor*, *general evaluation of the course*, *interest generated by the course*, *interest generated by the instructor*, *contribution to learning*, and *clarity of instruction*. Categories related to the instructor's personal traits, course assignments, scheduling issues, and student composition were all mentioned less frequently. It was found that students had a tendency to write more general comments about the course or the instructor as opposed to specific comments, and they referred more to interest in course content and learning as opposed to the more technical aspects of teaching.

Both positive and negative meaning units were identified for all 45 categories. Overall, more positive meaning units (59%) than negative ones (41%) were found. For example only five of the 18 most-mentioned categories (those mentioned by at least 5% of the students) had a mean score which was negative: *scope of material covered in the course*, *relevance of content taught*, *variation in instructional methods*, *the degree to which instruction is systematic and structured*, and *whether the course should be compulsory*. The most positive meaning units related to *evaluation of the teacher in general* (“He’s a great teacher”), *evaluation of the course in general* (“It was one of the best courses I have taken”), *how interesting the course was*, *how interesting the teacher was*, and *how much the student learned from the course*.

Consistency between written comments and student ratings

Pearson correlation coefficients were calculated between the dimensions of written comments and the dimensions of student ratings. The results are presented in [Table 2](#).

Findings revealed moderately positive correlations between four dimensions of the written comments (course content, teacher personal traits, teaching style, and general evaluation of instructor) and the three dimensions of student ratings (.36–.57). In general, contextual factors (scheduling issues and student composition) were unrelated to student ratings, while course assignments and the general evaluation of the course were only weakly correlated (.23–.36) to them. A similar pattern was detected regarding the correlations between the two general measures of student ratings and the various dimensions of written comments. It is worth noting that the general measure of student written comments was strongly correlated with all quantitative measures whether general or specific.

Discussion

Apparently not all students make the effort or feel the need to add written comments to the numerical ratings on course evaluation questionnaires. As in the studies by [Hardy \(2003\)](#)

Table 2
Pearson correlations between measures of the written comments and student ratings.

Focus	Measure of written comments	Student ratings				
		Course content and learning	Instructional management	Teacher–student relations	General evaluation (single item)	Rating mean score
Instructor	General evaluation	.505	.460	.521	.536	.545
	Teaching style	.521	.546	.491	.586	.568
	Personal traits	.403	.362	.452	.422	.445
Course	General evaluation	.355	.252	.228	.386	.323
	Content	.574	.390	.398	.586	.526
	Assignments	.252	.227	.313	.279	.287
Context	Scheduling issues	–.130	–.075	–.088	–.127	–.115
	Student composition	.128	.139	.041	.147	.117
General score		.677	.590	.609	.724	.697

Note: Correlation coefficients of absolute values greater or equal to .138 are statistically significant at $\alpha = .05$.

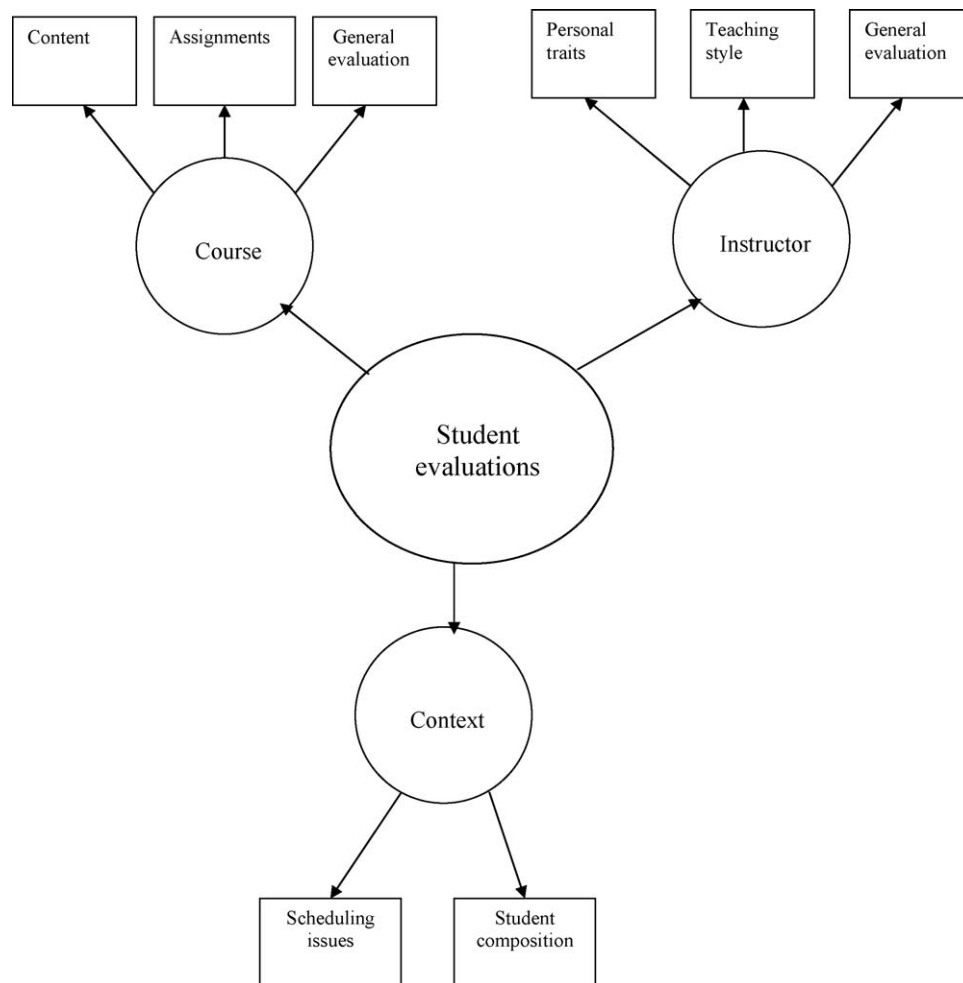


Fig. 1. Model depicting components of student written evaluations.

and Zimmaro and colleagues (2006), less than half of the students in the present study added some form of comment which ranged from a single word to several sentences. Not surprisingly, there was a tendency for students to write comments when the course evoked relatively strong reactions. However, more positive than negative comments were written overall. Similar findings have been reported by others (e.g., Braskamp, Ory, & Pieper, 1981 in Centra, 1993; Zimmaro et al., 2006).

Content analysis indicated that three major domains and eight primary content areas were addressed in students' comments: the course (content, assignments, and general), the instructor (teaching style, personal traits, and general) and the context of instruction (scheduling issues and student composition). The quantitative measure for evaluating faculty teaching in this study concentrated on course and instructor characteristics, generally ignoring contextual factors. Thus, one contribution of the analysis of students' comments has been to draw attention to the influence of contextual factors on students' course evaluations. Comments on contextual factors were less frequent than those that related to instructor and course factors, but when students did address these areas, they tended to be critical. One of the ramifications of these results is in relation to decisions made and actions taken on the basis of student feedback, namely that department chairs and promotion committees should be made aware of the possible negative influence of contextual factors when reviewing student feedback for evidence of good teaching.

Taken together the domains and primary content areas form a model, which depicts the principle components of a course that

play a factor in student evaluations. This model which emerged from analyzing the students' written comments is depicted in Fig. 1. Although our data indicate that contextual factors are secondary to instructor and course factors, further research is needed to understand the relative contribution of all the various components to students' evaluations. Moreover, research should investigate the relevance of the two contextual variables in other academic institutions and in other countries and an attempt should be made to identify other important contextual variables which influence students' perceptions of instruction.

One of the purposes of the present study was to see to what extent there was consistency between students' written comments and their numerical course ratings. Findings showed that the common dimensions of the two types of evaluation were highly correlated, whereas the unique dimensions of the written comments tended to be unrelated to either the general scores or the three dimensions of student ratings. These results can be interpreted as an indication of the validity of the written comments, but also once again point to their unique contribution to course evaluation. While student ratings which are commonly used for course evaluation provide a common base for judging instructional quality, they address limited aspects of teaching. As the findings presented here show, students' written comments can potentially address a broader range of variables and identify specific weaknesses and strengths of a particular course. It seems that a more comprehensive and valid picture of teaching can be depicted by applying a mixed-methods design in which quantitative ratings are complemented and supported by qualitative written comments.

General comments, relating to either the course or instructor, were found to be more frequent than students' comments in any specific area. This can be interpreted in several ways. First of all, it was probably easier and quicker for the students to write short general statements rather than to go into detail about various specific aspects of a course. If we take into account that the course evaluation questionnaire presented only general open-ended questions rather than questions that specified which course aspects should be addressed, this explanation has some validity. Moreover, course evaluation was administered during class time, which imposes a time constraint on writing detailed comments. Finally, since the closed-ended items which appear first on the evaluation questionnaire addressed specific aspects of the course, students may not have felt the need to repeat their ratings in words, feeling that a brief inclusive statement would suffice.

General comments are less useful to instructors if the purpose of evaluation is to improve instruction. If, like the research literature indicates, instructors actually prefer written comments to statistical summaries and feel that they can learn more from them, then measures should be taken to increase both the number of students who make their voice heard by writing comments on course evaluation forms and the quality of their comments. Svinicki (2001) has noted that many students do not have the skills required to provide detailed and useful feedback, but that these skills can be learned. She suggests that instructors supply feedback to students on the quality of their comments and encourage them to be more specific in what they write. Encouraging the students to write comments is a way of telling them that their voice is important and will likely result in more students writing comments as well.

Receiving detailed and numerous student comments are not sufficient in order for them to influence teaching. As pointed out by Lewis (2001), analyzing them and sorting them into meaningful categories is necessary in order for instructors to make sense of them. Imposing some structure on the comments should make them more comprehensible to instructors and could possibly yield better insights into their teaching. Moreover, in order for written feedback to be effective, it may be preferable to administer open-

ended questions at different times throughout a course rather than at the end. This way the instructor can take the comments into account with respect to the students writing them and not for some future group. Students who see that their instructors are attentive to their feedback will more likely be motivated to put an effort into writing meaningful comments.

Student feedback on instruction has been accused of not being serious and many faculty members are convinced that students are unqualified to provide valid assessments of teaching (e.g., Bonetti, 1994; Watchel, 1998). The results of the present study tend to contest this claim. When students' comments were specific, we found that they focused most frequently on four categories: the *interest* generated by the course and the instructor, the *contribution* of the course to learning and the *clarity* of teaching. In other words, their main concerns are not frivolous but show that they value learning and respond to a course with this goal in mind.

Despite the contribution of the present study to understanding students' written feedback on teaching, more research is needed. This study examined comments provided by students in response to two general questions after completing a closed-ended questionnaire. Future research should investigate students' comments which are given before they respond to closed-ended items and comments provided on questionnaires containing open-ended questions only. In addition, responses to open-ended questions which focus on specific aspects of the course, the teacher, and instruction need to be compared to responses to general open-ended evaluation questions. Furthermore, because of the importance attributed to students' written comments by instructors and their potential impact on the quality of instruction, studies in various types of institutions of higher learning, in other countries, and with different instruments will strengthen confidence in the validity of the information obtained using open-ended course evaluation measures.

Acknowledgment

We would like to thank Dr. Tzachi Ashkenazi for his contribution at an earlier stage of this study.

Appendix A

Content categories and examples of meaning units.

Focus	Primary: content area	Secondary content area	Example
Instructor	General Teaching	General evaluation Interest Clarity Structure Classroom environment Diversity of methods Contribution to learning Activating students Efficient time use Combining theory and practice Use of teaching aids Pace Supervision quality	He is a great teacher. She presents the material in a very interesting way. She explained everything very clearly. He should present the material in a more structured and organized manner. She creates a good environment in class. He should include movie clips and visits to the archives. He helps improve the way we think. Discussions were good. Class time was often wasted. She should emphasize more practice and less theory. The articles to be read should be put on the internet. She talks too fast. She doesn't know how to provide individual supervision when we meet her during her office hours.
	Personal traits	Conduct towards students Flexibility and consideration Professional competence Commitment Emotional support Ability of influence students Responsiveness Sense of humor	He treats the class badly. She was always willing to meet with students even outside her office hours. She has a great wealth of knowledge and is very professional. I appreciate how much the instructor is devoted to his work. Without her encouragement I would have dropped out of this course. I didn't expect to enjoy this course so much but because of the teacher, I really began to like this subject. Whenever we did not understand something, she was always patient with us. He has a great sense of humor.
Course	General	General evaluation	This is a very good course

Appendix A (Continued)

Focus	Primary: content area	Secondary content area	Example
	Content	Interest	The course was really interesting.
		Contribution to learning	I finally learned how to write a paper.
		Scope	Much of the material taught in this course I already learned in other courses.
		Relevance	She should teach probability like we have to teach it in high school.
		Compulsory	It is good that this course is required.
		Importance	The material in this course is important for teachers.
		Difficulty	Course material was very difficult and very abstract.
		Match with syllabus	We never got to all the topics that we were supposed to cover.
		Combining theory and practice	The course material was too theoretical.
	Assignments	Quantity	We need more practice, so we should get more homework.
		Relevance	He should give a paper to do at the end of the course and not a test.
		Difficulty	He gives three assignments and a test and every thing is complex and difficult.
		Clarity	We were given very detailed instructions regarding our final assignments.
		Contribution to learning	The homework helped me understand course material.
		Timing	Assignments should be given throughout the term and not just at the end.
Context	Scheduling issues	Length of course	This course was too long and tiring.
		Hour of class meetings	A course like this should be offered earlier and not at the end of the day.
		Sequence in training program	This course should not be given during the third year of study, but sooner.
	Student composition	Class size	The class was too crowded.
		Heterogeneity	Some of the problems were because students had different prior knowledge.
		Discipline problems	Disturbances made by students during class were intolerable.
		Overall academic level	We were held up all the time by certain students who did not understand anything and kept asking questions.
		Overall motivation	Some of the students were just not interested in learning.

Appendix B

Factor analysis results for student ratings (N = 198).

Item	Course content and learning	Instructional management	Teacher–student relations
Course content is stimulating.	.988	.009	.121
This course can contribute to my profession work in the future.	.773	.062	.011
In this course, I am given the opportunity to demonstrate my knowledge and competence.	.704	.053	.276
The content of the course is presented in an interesting way.	.621	.254	.149
Discussions in class contribute to my understanding of course material.	.585	.290	.170
Assignments contribute to my understanding of course material.	.582	.359	.066
In this course there are many opportunities for students to be active.	.582	.016	.349
The instructor defines course requirements clearly.	.062	.955	.016
The instructor teaches the course according to the syllabus.	.119	.782	.076
The instructor provides clear instructions for all assignments.	.003	.778	.186
Class time is efficiently used.	.413	.582	.056
The instructor encourages an exchange of ideas in class.	.179	.164	.713
The instructor is tolerant of students' views that differ from his/her own.	.036	.383	.623
The instructor is not considerate of individual differences among students (R).	.081	.009	.578
The instructor creates a pleasant learning environment.	.245	.376	.477
Course assignments are too difficult (R).	.026	.047	.368
Eigenvalue	8.39	7.51	5.40
% Explained variance	61.66%	5.94%	4.51%
Cronbach's α	.95	.91	.83

References

- Bonetti, S. (1994). On the use of student questionnaires. *Higher Education Review*, 26(3), 57–64.
- Cashin, W. E. (1995). *Student ratings of teaching: The research revisited*, IDEA paper no. 32. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
- Cashin, W. E. (1998). *Student ratings of teaching: a summary of the research*, IDEA paper no. 22. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
- Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Feldman, K. A. (1976). A superior college teacher from the student's view. *Research in Higher Education*, 5, 243–288.
- Feldman, K. A. (1989). Instructional effectiveness of college teachers as judged by teachers themselves, current and former students, colleagues, administrators, and external (neutral) observers. *Research in Higher Education*, 30, 137–189.
- Feldman, K. A. (1993). College student views of male and female college teachers: Part II—Evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, 34, 151–211.
- Hammond, L., Taylor, J., & McMenamin, P. (2003). Value of structured participant evaluation questionnaire in the development of a surgical education program. *Australian & New Zealand Journal of Obstetrics & Gynecology*, 43(2), 115–118.
- Hardy, N. (2003). Online ratings: fact and fiction. *New Directions for Teaching and Learning*, 96, 31–38.
- Hodges, L. C., & Stanton, K. (2007). Translating comments on student evaluations into the language of learning. *Innovations in Higher Education*, 31, 279–286.
- Johnson, T. D. (2003). Online student ratings: Will students respond? *New Directions for Teaching and Learning*, 96, 49–59.
- Lewis, K. G. (2001). Making sense of student written comments. *New Directions for Teaching & Learning*, 87, 25–32.
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues and directions for future research [Entire issue]. *International Journal of Educational Research* 11(3).

- McGee, R. (1995). Faculty evaluation procedures in 11 western community colleges. *Community College Journal of Research and Practice*, 19(4), 341–348.
- Nasser, F., & Fresko, B. (2002). Faculty views on student evaluation of college teaching. *Assessment and Evaluation in Higher Education*, 27(2), 187–198.
- Nasser, F., & Hagtvet, K. A. (2006). Multilevel analysis of the effects of student and instructor/course characteristics on student ratings. *Research in Higher Education*, 47(5), 559–590.
- Obenchain, K. M., Abernathy, T. V., & Wiest, L. R. (2005). The reliability of students' ratings of faculty teaching effectiveness. *College Teaching*, 49(3), 100–104.
- Oliver, B., Tucker, B., & Pegden (2007, 11–13 July). *An investigation into student comment behaviors: Who comments, what do they say, and do anonymous students behave badly?* Paper presented at the Australian Universities Quality Forum, Hobart.
- Ory, J. C., Braskamp, L. A., & Pieper, D. M. (1980). Congruency of student evaluative information collected by three methods. *Journal of Educational Psychology*, 72(2), 181–185.
- Renaud, R. D., & Murray, H. G. (2005). Factorial validity of student ratings of instruction. *Research in Higher Education*, 46(8), 929–953.
- Schmelkin-Pedhazur, L., Spencer, K. J., & Gellman, E. S. (1997). Faculty perspectives on course and teacher evaluations. *Research in Higher Education*, 49(3), 497–511.
- Sheehan, E. P., & DuPrey, T. (1999). Student evaluations of university teaching. *Journal of Instructional Psychology*, 26(3), 188–193.
- Smith, K., & Welicker-Pollack, M. (2008). What can they say about my teaching? Teacher educators' attitudes to standardized student evaluation of teaching. *European of Teacher Education*, 31(2), 203–214.
- Sorenson, D. L., & Reiner, C. (2003). Charting the uncharted seas of online student ratings of instruction. *New Directions for Teaching and Learning*, 96(1), 24.
- Svinicki, M. D. (2001). Encouraging your students to give feedback. *New Directions for Teaching and Learning*, 87, 17–24.
- Symons, R. L. (2006, November). *Listening to the student voice at the University of Sydney: closing the loop in the quality enhancement and improvement cycle* Paper presented at the 2006 Australian Association for Institutional Research (AAIR) Forum, Coffs Harbor, NSW.
- Theall, M., & Franklin, J. (1991). Using student ratings for teaching improvement. *New Directions for Teaching and Learning*, 48, 83–96.
- Watchel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, 23(2), 191–211.
- Zimmaro, D. M., Gaede, C. S., Heikes, E. J., Shim, M. P. & Lewis, K.G. (2006). A study of students' written course evaluation comments at a public university. <http://www.u-texas.edu/academic/mec/publication/pdf/fulltext/courseevalcomments.pdf>.