

Provost's Advisory Committee on Teaching and Learning (PACTL) Report to Senate

REVIEW OF NEW UNIVERSITY-WIDE TEACHING EVALUATION FORM

Introduction

At its June, 1996 meeting, Senate approved a new Instructor and Course Evaluation form (see Exhibit 1) for university-wide use beginning in the 1996-97 academic year. The distinctive features of the new teaching evaluation form include numerical ratings of a broad range of teacher and course characteristics, written comments from students on instructor and course, assessment of student characteristics such as class attendance and expected grade, and use of a 7-point poor-outstanding rating scale rather than the traditional 5-point agree-disagree scale. The teacher and course characteristics included on the new teaching evaluation form were selected according to the following criteria: (1) observable by students, (2) under the control of the instructor, (3) applicable to all or nearly all

Table 1 shows the percentage of students who agreed, disagreed, or were undecided for each of the 11 survey items. For purposes of this table, "agree" and "strongly agree" responses have been combined into one category, as have "disagree" and "strongly disagree" responses. It may be noted that students were generally very positive about the new teaching evaluation form. Approximately 85% of respondents said that the new evaluation form assesses a sufficiently wide range of characteristics, whereas 81% considered the form to be applicable to the style of teaching to which they were accustomed, 92% thought the written comments section of the evaluation form was useful, 77% believed the new 7-point rating scale was more appropriate than the previous 5-point scale, and 60% agreed that, all things considered, the new evaluation form was superior to previous forms with which they were familiar. Students were strongly opposed to deleting the written comments section or making it optional for individual faculty members (90% and 84% disagreement respectively). On the other hand, only 20% of respondents thought that sufficient weight was placed on student evaluation of teaching in faculty personnel decisions at UWO, only 14% were aware that teaching evaluations are published on the UWO Web site, and only 4% said they used published evaluations in selecting courses. The latter two results may be due to the fact that last year's teaching evaluations were not available on the Internet until after most students had already registered for courses.

Table 2 summarizes students' supplementary comments regarding the new evaluation form. A total of 203 individual comments were received, which could be grouped into 24 categories. Only those categories with a response frequency of 10 or higher are reported in Table

comments were actually abusive or obscene, and how many were inappropriate for other reasons (for example, they were silly or childish or unfounded). Also, contrary to the initial expectation that reports of abusive written comments would be more frequent for female than for male faculty members, breakdown of survey responses according to gender indicated approximately equal frequencies for females and males (27% and 29 % respectively). Finally, it should be noted that although abusive comments were reported by a sizeable minority of respondents, the majority of faculty believed that written comments provided beneficial

Table 6 shows correlations between various teacher and course characteristics and instructor mean ratings on the overall effectiveness item (#19) of the new teaching evaluation form. These results are based on data for 103 psychology classes taught in the 1996-97 academic year. It may be noted that, on average, overall teacher effectiveness ratings were significantly higher for senior faculty (eg., full professors) than for junior faculty (eg., instructors); significantly higher for senior (eg., Year 4) courses than for junior (eg., Year 1) courses; significantly higher for courses with higher attendance levels on the day of evaluation; significantly higher for courses with higher reported attendance for the course as a whole (student information item #1); significantly higher for courses with higher mean expected final grades (student information item #2); and significantly higher for courses with higher mean ratings of initial level of student interest (student information item # 4). On the other hand, overall teaching effectiveness ratings did not differ significantly, on average, for female vs. male faculty members, for Fall Term vs. Spring Term administration of the teaching evaluation form, for small vs. large classes, or for required vs. optional courses. Results similar but not identical to those in Table 6 were obtained when the same analyses were repeated for two other departments with large undergraduate enrollments, namely English (N=160 classes) and Mathematics (N=85 classes). For example, course status (required vs. optional) correlated significantly ($r = -.34$) with instructor mean ratings in Mathematics, but not in English or Psychology, whereas instructor rank correlated significantly with ratings in Psychology but not in English or Mathematics.

For most of the instructor and course characteristics listed in Table 6, it is difficult to know whether a significant correlation with instructor mean rating should be interpreted as evidence of "bias" or "error" in student evaluation of teaching, or as a valid reflection of factors that contribute to effective teaching. For example, a significant positive correlation between instructor rank and teacher rating could be interpreted to mean that students are biased against younger teachers, or alternatively, as a tendency for increased teaching effectiveness resulting from age and experience to be validly reflected in student ratings. Similarly, a positive correlation between mean expected grade and mean teacher rating could reflect a tendency for students to "reward" lenient-grading teachers with high ratings, or alternatively, could mean that students actually learn more in courses taught by more effective teachers, and this higher level of learning is reflected both in higher grades and in higher ratings of the instructor. Even in cases where correlation with instructor and course characteristics can be unambiguously interpreted as "bias" or "error", it is important to bear in mind that because of intercorrelation of bias factors, all sources of bias in combination typically account for only 10 to 15% of the total variance in instructor mean ratings. Also, it is possible to eliminate the impact of most sources of bias through the use of statistical adjustments or separate norm groups for different types of courses.

Table 7 compares the frequency distribution of psychology instructors' mean ratings on the overall effectiveness item of the new teaching evaluation form (1996-97 data) to the distribution of ratings on the corresponding item of the istriPily t ef45% of ths acove2.55ratin8s on ra-3.7(asesv6(ra-3sTJ8(..7(asesn6(ra-3)TJ7(w.7(ases

student ratings, whereas feedback supplemented by expert consultation produced a much larger (.40) gain in end-of-term ratings.

3. Longitudinal analyses of mean teacher ratings over periods of 3 to 25 years following the introduction of student evaluation of teaching in a given academic unit have generally (but not always) found gradual improvement in perceived quality of teaching across years.

4. Undesirable educational practices such as grade inflation, watering down of academic standards, and entrenchment of traditional "hierarchical" methods of teaching are sometimes attributed to student evaluation of teaching, but there appears to be no systematic empirical evidence to support these claims.

Reference

Murray, H. G. (1997). Does evaluation of teaching lead to improvement of teaching? *International Journal of Academic Development*, 2 , 8-23.

Table 1

Results of Student Survey

Questionnaire Item (paraphrased) and Percent Responses

1. The new teaching evaluation form assesses a sufficiently wide range of instructor and course characteristics.

Agree: 84.8

Disagree: 6.2

Undecided: 9.0

2. The items on the new evaluation form are generally applicable to the style of teaching to which I am accustomed.

Agree: 80.9

Disagree: 3.9

Undecided: 15.2

3. The separate course evaluation item is a useful feature of the new form.

Agree: 85.4

Disagree: 3.3

Undecided: 11.3

4. The written comments section allows me to comment on aspects of teaching not covered by numerical items.

Agree: 91.8

Disagree: 1.7

Undecided: 6.5

Undecided: 9.2

6. The written comments section should be deleted from the new evaluation form.

Agree: 3.3

Disagree: 90.0

Undecided: 6.7

7. I was aware that instructor and course evaluations are published on the UWO Web site.

Agree: 14.4

Disagree: 78.5

Undecided: 7.1

8. I used published teaching evaluations on the UWO Web site in selecting courses for this year.

Agree: 4.4

Disagree: 89.4

Undecided: 5.2

9. The 7-point rating scale on the new evaluation form is more appropriate for evaluation of teaching than the 5-point scale on previous evaluation forms.

Agree: 76.7

Disagree: 7.1

Undecided: 16.2

10. All things considered, the new teaching evaluation form is superior to other evaluation forms with which I am familiar.

Agree: 59.7

Disagree: 2.9

Undecided: 37.4

11. Sufficient weight or emphasis is given to student evaluation of teaching in decisions on faculty salary, tenure, and promotion at UWO.

Agree: 20.4

Disagree: 38.7

Undecided: 40.9

Table 2

Student Survey: Content Analysis of Supplementary Comments

Category Frequency

Student evaluation of teaching is not taken seriously in personnel decisions or improvement of teaching.	51
--	----

There should be more items evaluating the course as opposed to the instructor (e.g., course quality, reading materials, work load, grading system).	26
---	----

Students are not made aware of the uses or purposes of student evaluation of teaching.	22
Some items on the new evaluation form are inappropriate (i.e., unclear, inapplicable, or too general).	20
The written comments sections of the new evaluation form are useful or valuable.	14
There should be increased opportunities for student evaluation of teaching (e.g., more time available, mid-term evaluation, alternative date for those who missed class).	10

Note: Results are reported only for comment categories with response frequencies of 10 or higher.

Table 3

Results of Faculty Survey

Questionnaire Item (paraphrased) and Percent Responses

1. Results of student information items should be made available to faculty.

Agree: 91.0
 Disagree: 1.5
 Undecided: 7.5

2. Instructor ratings should be adjusted to account for factors such as expected grade and percent attendance.

Agree: 70.3
 Disagree: 14.9
 Undecided: 14.8

3. The new teaching evaluation form assesses a sufficiently wide range of instructor characteristics.

Agree: 73.4
 Disagree: 15.0
 Undecided: 11.6

4. The items on the new evaluation form are generally applicable to my style of teaching.

Agree: 72.1
 Disagree: 16.6
 Undecided: 11.3

5. The separate course evaluation item is a useful feature of the new form.

Agree: 72.1
Disagree: 13.2
Undecided: 14.7

6. The new 7-point rating scale provides more useful data for salary, promotion, and tenure decisions than the previous 5-point scales.

Agree: 42.0
Disagree: 30.7
Undecided: 27.3

7. Designation of items as "not applicable" should be decided in advance by the instructor rather than by individual students.

Agree: 53.4
Disagree: 24.4
Undecided: 22.1

8. Written comments provide a useful supplement to numerical ratings.

Agree: 72.3
Disagree: 19.8
Undecided: 7.9

9. Written comments provide useful feedback for improvement of teaching.

Agree: 59.8
Disagree: 25.8
Undecided: 14.4

10. I received abusive, obscene, or inappropriate written comments in last year's teaching evaluation.

Agree: 28.2
Disagree: 65.0
Undecided: 6.8

11. Written comments should be used as feedback to the instructor but not for decisions on faculty salary, promotion, and tenure.

Agree: 74.4
Disagree: 18.4
Undecided: 7.1

12. Individual faculty members should decide whether students complete the written comments section of the evaluation form.

Agree: 22.0
Disagree: 64.9
Undecided: 13.1

13. The written comments section should be deleted from the new evaluation form.

Agree: 16.1
Disagree: 72.0

Undecided: 11.9

14. Faculty members should receive the results of annual teaching evaluations by June 1.

Agree: 90.4

Disagree: 5.0

Undecided: 4.6

15. Feedback to faculty should include teaching evaluation norms for different departments, course types, and class sizes.

Agree: 83.7

Disagree: 6.1

Undecided: 10.2

16. The new teaching evaluation form provides useful feedback for improvement of teaching and courses.

Agree: 64.2

Disagree: 16.4

Undecided: 19.4

17. The data provided by the new teaching evaluation form are suitable for use in salary, promotion, and tenure decisions.

Agree: 69.2

Disagree: 18.5

Undecided: 12.1

18. All things considered, the new teaching evaluation form is superior to evaluation forms used previously in my department.

Agree: 43.8

Disagree: 25.5

Undecided: 30.7

Table 4

Faculty Survey: Content Analysis of Supplementary Comments

Category Frequency

The new evaluation form focuses too much on instructor "classroom performance", and not enough on course quality and amount learned by students.	30
Some items on the new evaluation form are inappropriate for certain types of courses.	27
Feedback to instructors should occur earlier, especially for first-term courses.	26
Written comments from students are sometimes abusive,	19

rude, irresponsible, or contradictory.

Students should be required to sign their names on teaching evaluation forms. 17

Student evaluation of teaching is biased, invalid, or influenced too much by teacher "popularity". 13

The new teaching evaluation form favours a transmissive, lecture-style method of teaching, and thus is not suitable for use in all departments or all types of courses. 10

Note: Results are reported only for comment categories with response frequency of 10 or higher.

Table 5

Reliability of New Teaching Evaluation Form

(Sample: 116 psychology classes in 1995-96 academic year, 103 psychology classes in 1996-97 academic year; unit of analysis: class mean ratings)

Reliability Test Coefficient	Reliability
Mean interrater reliability (split-half) for instructor evaluation items #5 to 19 (N= 103)	.84 *
Mean intercorrelation of instructor evaluation items # 5 to 18 (N=103)	.67*
Mean correlation of instructor evaluation items # 5 to 18 with overall evaluation item # 19 (N=103)	.80 *
Correlation of overall instructor rating (item #19) with mean rating for all other items combined (N=103)	.95 *

Correlation of overall instructor rating (item # 19) .62 *

across different courses taught by same instructor
in same academic year (N= 87)

Correlation of mean instructor rating on old vs.
new teaching evaluation forms (all items combined):
Same Course, Same d vs.

* Statistically significant at .05 level

Table 7

Distribution of Overall Instructor Ratings for Old vs. New Evaluation Forms

(Sample: 116 psychology classes in 1995-96 academic year, 103 psychology classes 1996-97 academic year; unit of analysis: class mean ratings)

Old Form /New Form

(5-point scale, item #10, N=116) (7-point scale, item #19, N=103)