AN ESSAY
ON
STUDENT EVALUATION OF TEACHING

Prepared by
P. A. Sullivan
for

TEACHING METHODS AND RESOURCES COMMITTEE
FACULTY OF APPLIED SCIENCE AND ENGINEERING
UNIVERSITY OF TORONTO

September 1975

CONTENTS

Note by author: I recently edited this essay for electronic storage and transmission, and I make two points for the 21st Century reader. First, as will be evident from the language, the faculty and students for whom this essay was prepared were overwhelmingly— if not exclusively—male, a state of affairs which has since radically changed. However, I consider it dishonest to revise the language to reflect current sensitivities. The second point is that many of the recommendations for policy have long since been implemented, so that much of Sections 6 and 7 and Appendix C is not now relevant. However the issues raised in the survey of the literature at the time of original writing remain valid.

P.A. Sullivan
December 22, 2004

1. INTRODUCTION

A basic task of the Teaching Methods and Resources Committee is to encourage the adoption of effective methods of recognizing and rewarding teaching excellence in this Faculty. Of major interest is the role of student opinion in the process, and the purpose of this essay is to provide a basis for discussion and development of a suitable policy. In general, student participation can be solicited through formal teaching evaluation questionnaires (TEQ), signed or unsigned free comments, and interviews. The questionnaires, which may be administered by staff or students, usually seek responses on the adequacy of lecturing, laboratories, tutorials and any aids used, as well as on the relevance of the course itself. The results may be used to provide the individual professor with information on the quality of his teaching and the usefulness of the course, in promotion and tenure deliberations, and possibly as a guide to course selection by the students. In this Faculty, questionnaires and unsigned free comments are the only methods by which information is formally sought from students.

There arc three reasons why a review is now appropriate. Firstly, there has recently been widespread discussion in this and other university communities on the adequacy of university teaching. Indeed, there are many who view the current university system as far too heavily oriented towards research at the expense of teaching, and consequently there has been considerable pressure from some sources to make formal evaluation procedures compulsory in order to rectify the perceived problem. University of Toronto policy on appointments, promotion, and tenure now reflects this concern, and both signed student opinions and results from questionnaires are cited as items to be considered at the appropriate proceedings (Ref. 1). Secondly, there is a very considerable body of research now available on student evaluation of teaching and this, as will be seen, provides very interesting perspectives on the subject. Typically a review by Costin, Greenough and Menges published in 1971 (Ref. 2) cites well over a hundred empirical studies. Indeed, it seems that much of the current debate on the role of the TEQ takes place in ignorance of this work. Consequently this debate often consists of little more than statements of a range of personal and largely unsubstantiated opinions. The third reason is that course evaluation procedures have been widely used in this Faculty for some time; in one department such use began in 1966. These procedures have been developed on an individual and largely informal basis. There is evidence that TEQ results have played a significant role in this Faculty in promotion and tenure deliberations, and it is clear that they will become increasingly important. Consequently, it is desirable that their adequacy be reviewed, especially in the light of the available research information.

Several sources of information have been used in preparing this document. Department heads were interviewed in order to ascertain the range of current practices, and a brief questionnaire was circulated amongst the Faculty's academic staff in order to assess attitudes. Discussions have been held with several professionals in the fields of psychology and education, and extensive use is made of reported empirical studies. The review by Costin, Greenough and Menges (Ref. 2) together with a more recent one by Murray (Ref. 3) have provided most of the source material. Finally, this essay has been checked for factual correctness by a professional in the field (Ref. 25).

It should be noted that the interested layman faces a basic difficulty in interpreting the literature on teaching evaluations. Apart from the fact that an appropriate training in experimental psychology is required to fully understand the results of the experiments, teaching evaluations research is prone to the vagueness, controversies and dogmatism that is endemic to the social sciences. Furthermore, the difficulty in obtaining adequate controls when undertaking an

experiment on a subject as complex as the interaction between a lecturer and his students implies that any empirical study has to be repeated, possibly many times, before the conclusions can be accepted as having any useful generality. Accordingly, the author's approach has been to accept propositions as valid only if they have been backed up by several independent experiments. If not, it was assumed that nothing had been established. This approach, which has been suggested by those disturbed by the pseudo-scientific quality of much of the research in the social sciences (Ref. 4), turns out to be a useful test. Hence judgments based on the results of experiments have been taken as valid if there seems to be widespread agreement in the professional literature as a direct result of these experiments. Some typical reported correlations of traits and factors have been given below to provide some insight to these judgments, but the primary bases were the judgments themselves.

2. A STATEMENT OF THE PROBLEM

The general question of effective promotion of good teaching has been the subject of discussion among concerned faculty members for much longer than might be inferred by readers of Varsity. The issue touches upon the basic question of the special nature of a university institution. Contrary to the assertions of certain student activists and some politicians, many professors do not see teaching and research as separable activities which can individually compete for attention, the one to the detriment of the other. It is commonly believed that, in any university worth the name, the two are so intermingled that they mutually reinforce each other and give a special flavour to a man's understanding of his subject which is not usually attainable in institutions devoted solely to either teaching or research. Indeed, there is a traditional view that they are so closely related that it is unnecessary to evaluate research and teaching separately (Ref. 3, p. 24), so that the classical criterion of excellence in scholarship is still widely seen as being all that is necessary, with teaching quality being an essentially private matter (Ref. 5). Hexter's article, "Publish or Perish—a Defense" (Ref. 6) puts this view very strongly.

Empirical studies give important insights into this problem. The question can be posed as follows: Is there any correlation between various possible measures of teaching performance, such as peer opinions and student evaluations, and the accepted measures of research performance? If such correlations were highly positive, then it would support the traditionalist view that independent evaluation of teaching is unnecessary, whereas a highly negative correlation tends to support what might be colloquially described as the student activist view. Studies have been reported in which both weakly negative and weakly positive correlations have been obtained (Ref. 2), but in general, the consensus is that performance in teaching is effectively unrelated to either quantity or quality of research. Typically, Hayes (Ref. 7) collected data on 35 professors in 17 departments at a well-known American university. Research criteria used were publication rate, grant status and opinions of department chairmen. Teaching ratings were obtained from students and departmental heads. Of the six possible correlations, only one was reported as statistically significant; this was the strongly positive correlation between department head ratings of teaching ability and their rating of research competence. This however was thought to be a halo effect.[1] McDaniel and Feldhusen (Ref. 8) computed 72 correlations between individual items on an 18 item TEQ with various

[1]In psychological practice, the term *halo effect* means the general tendency of an observer of an individual to associate two different unrelated properties. A simple example is a student giving a good teaching rating to a man he likes personally.

measures of research performance. Only 21 were statistically significant and the largest of these was only 0.33. There was significant difference between the ratings of professors with research grants and those without grants. Apparently there was a weak but persistent tendency for evaluations to correlate negatively with first authorship of books or journal articles but positively with second authorship.

In general, there is widespread agreement in the literature that sole reliance on the research criterion in university reward structures does not adequately recognize and encourage good teaching, so that specially developed procedures are required. E. F. Sheffield, Professor of Higher Education at the University of Toronto, has made this point very effectively in a recent study of teaching in Canadian institutions (Ref. 9). Owing to its somewhat unusual nature and the fascinating insights obtained, this study merits special comment. In response to a request from Professor Sheffield, roughly a thousand alumni nominated professors they had thought were excellent teachers, and gave reasons for their selection. From these, twenty-three of the most frequently nominated professors were asked to write essays on their own personal view of teaching. The published study includes the essays, summaries of the comments made by alumni about the selected professors, and Sheffield's own penetrating analysis. Several references are to be made below to this study; but of immediate interest is one of his major conclusions. Teaching was widely identified as requiring thorough preparation and much hard work, but it is also made more effective through research and field experience. He goes on to say (Ref. 9, p.216)

> Here is raised the crucial question of priorities. How is a professor to divide his time and effort? For most, the competing activities are teaching, research, writing, administration, public service. The requirements of his department will limit a professor's choice among these activities. Within such limits, however, his personal preferences will influence the distribution of his time. But perhaps most influential will be the extent to which each type of activity is rewarded—in terms of rank, salary, and professional recognition. A major challenge to the university system is to discover how to reward teaching in proportion to the effort put into it. Then teaching could compete on equal terms. Then professors with the desire to emphasize teaching would have the additional motivation to give it their preferred priority.

3. DEFINITION OF EFFECTIVE TEACHING AND ITS MEASUREMENT

Any attempt to construct a rating instrument or procedure implies that it is possible to agree on what constitutes "good teaching." This in turn might allow definition of quantifiable criteria. There are, of course, many who would object at the outset to even this basic step. Objections frequently heard are usually variations on the theme that teaching is a subjective activity which is comparable in many ways to an art form. Hence, it is argued, attempting to recognize good teaching by using quantifiable criteria is dangerous and naive. No doubt much of teaching has to be so described. Sheffield states

> It is complex, subtle, varying and dependent more on personality attitudes and human relationships than methods. (Ref. 9, p. 200).

But the purpose of teaching is learning by the students who are exposed to the teacher, and

> Unless the students have achieved the learning that was the object of the teaching, it would be odd to protest that the teaching had been efficient (Ref. 10, p. 11).

One of Sheffield's major conclusions is essentially the same:

> Teaching is effective only when learning is effective; what the student does is the important thing (Ref. 9, p. 215).

Now it is clear that many and possibly most aspects of learning can be measured in some objective sense. This applies to the three types of learning commonly identified: simple acquisition of facts, cultivation of skills, and development of an appreciation of such norms of in tellectual behaviour as critical reasoning (Ref. 10, p. 6).

The extent to which this can be done obviously depends on the nature of the subject, and in some areas this approach would probably be naive. For example, the problem might be expected to be especially acute in the humanities. One can easily conceive quantifiable indicators of competence in English grammar, but assessment of the ability to appreciate literature or art may be a different story. However, as engineers, we are fortunate that this thorny issue can be largely side-stepped. The overall objective of an engineering education is to develop a certain level of professional competence in the student, and this implies achieving certain objectives which are usually specific, relatively easy to define and free of controversy. Furthermore, the objectives can normally be translated into the development of a problem-solving ability, and this applies to both analytic subjects such as theoretical physics and synthetic subjects such as mechanical design. Of course there will frequently be controversy amongst engineering professors over choice of courses and the range of material to be covered in particular courses but this is to be distinguished from the basic idea proposed here.

It follows from this approach that one working definition of a basic aspect of teaching is "any procedure which manipulates the student's environment in such a way as to bring about a behavioural change" (Ref. 3. p. 1). For engineering students at least, good teaching can be measured to a large extent in terms of improvements in problem-solving ability. This definition suggests a measurement procedure; in principle it is only necessary to measure a student's improvement in this ability by means of appropriately designed examinations. A better teacher should, on average, inculcate more learning as measured by such tests.

Although this behavioural modification approach leads to an attractively simple and clear-cut method of assessment of teaching quality it must be realized that many would regard it as quite inadequate or at least incomplete. For example, Sheffield's essay writers are generally agreed that the teacher's most important role is to stimulate students to become active learners on their own. Typically, Professor Jolliffe of Queen's University, an engineer, says

> Surely this must be the teacher's prime function: to strike sparks, to build fires, so that the student's own enthusiasm will carry him through those dreary stretches that still exist in all learning (Ref. 9, p. 199).

In other words, as Murray has observed (Ref. 25), it may be possible for a teacher to create a persisting desire to learn, even though he may not be successful in producing a high level of comprehension of a particular course. Most, if not all, would agree that, for a university graduate, a positive attitude towards learning is a more important longer term objective than mastery of specific material. However, it would seem that the two aspects of teaching effectiveness just discussed are not unrelated, and that the situation in which a student developed an enthusiasm for

learning but attained poor levels of achievement in the types of learning described above would perhaps be somewhat unusual.

Apart from the possible conceptual difficulties associated with the behavioural modification approach to assessment of teaching, there are serious practical problems involved in its implementation. One would have to design standardized pre-course and post-course tests which could not be allowed to change very much in style, level or content from year to year. In elementary courses this might be possible, but in advanced level courses the objectives are often very flexibly defined, and the selection of material suitable for examination problems is often limited, so that standardized tests become impractical. Other questions raised are problems such as the comparison or a 10% level of achievement in senior years with a 50% achievement in first-year courses, isolating the effects of other courses and other teachers on the student's ability to absorb material, and constraints imposed by organizational factors. In most universities these constraints are such that they may prevent a professor from displaying his particular set of abilities in the way that suits him. Thus, as Falk and Dow (Ref. 10, p. 7) have noted:

> The course may have been designed by other people, parts of it may be taught by other lecturers, as may tutoring and laboratory work, and the teacher we wish to assess may not have authority to prescribe the methods of teaching. He may not design the whole examination nor be the sole assessor. It is therefore easier to say whether a course is well or badly taught than it is to determine for what each member of a team may be properly held accountable.

These difficulties are such that the behavioural modification or direct method of assessment can be ruled out for anything but the occasional experiment.

Owing to the technical obstacles confronting the use of the direct method, and because of widespread doubt about completeness of the assessment that would be obtained, it is universal practice to resort to other methods which might be collectively described as *indirect*. Such methods usually involve soliciting opinions from various groups; and students, academic peers, alumni, administrators have all been used on occasion. The TEQ given to students currently taking the course is the method most frequently used, since it is the easiest to administer and its proponents argue that students have the most immediate interest. Another possible indirect method is a classroom visitation or inspectorial system, but this is almost universally rejected as unacceptable both on technical grounds and because it inevitably creates morale problems (Ref. 3, p. 3).

Although great care may be exercised in the design and administration of TEQs they are, nevertheless, sampling procedures on judgments of competence made independently of any objective measure of teaching effectiveness in either of the two senses discussed above. Accordingly, a fundamental question raised by the use of such procedures is whether they correlate with more objective measures of student achievement (Ref. 3. p. 1). Most of the frequently voiced objections to the use of TEQs relate to various aspects of this question. Typically, it is often argued that TEQs will reduce teaching to “popularity contests”, or that lecturers who grade hard will tend to receive poor ratings from students, that poorer students will tend to give poor ratings and so on. The bulk of the research done to date relates to attempts to answer this fundamental question, and, as will be seen, it is a very complex task. A second basic problem is related to a difficulty perceived in the use of direct methods. It is necessary to ascertain the extent to which TEQs, when used in conjunction with a particular course, can effectively compensate for the influences beyond the lecturer's control.

4. THE DESIGN AND ADMINISTRATION OF QUESTIONNAIRES

The details of the design of a particular questionnaire depend on the intended uses. For example, if the primary objective is promotion and tenure deliberations, then a question on the usefulness of a text book may unfairly penalize the teacher of an advanced course for which no adequate text is available. Since promotion and tenure considerations are made at Faculty level, there needs to be uniformity in design and administrative procedures for questionnaires put to this purpose. On the other hand, departmental level control seems desirable for those parts of a questionnaire relating to course content, adequacy of laboratories and similar matters.

A systematic procedure commonly used for development of a suit-able TEQ is as follows. The specific objectives of the questionnaire should be agreed upon by a representative group, normally a joint committee of staff and students. The next step is to obtain an appropriate consensus on the factors that professors and students perceive as being the most important. This enables a set of statements to be selected which form the basis of the questionnaire. These statements should be simple,unemotionally worded and related to a single trait or attribute. Finally the questionnaire should be refined during an appropriate trial period in order to eliminate ambiguities, and items which are redundant and do not discriminate. This ensures that the questionnaire is as sensitive as possible.

Empirical studies on the selection of attributes for questionnaires have been reported and the results are very interesting. Perry (Ref. 11), for example, asked a total of over 6,000 students at four American universities to rank in order of importance a total of sixty possible attributes previously selected from over 1,300 submitted to a panel of judges. The results of this ranking, which are given in AppendixA, showed remarkable unanimity and greater maturity than many professorswould expect. Item 35, "being well prepared for class", was given first place in all four universities, and "establishing a sincere interest in the subject" (Item 27) was second. Demonstrating a "comprehensive knowledge of the subiect" (Item 20) was also ranked high. Use of visual aids, and such quirks as "having irritating mannerisms" appeared near the bottom of the list, and administrative items such as "explaining grading procedures" appeared below the middle rank. Nadeau (Ref. 15) quotes a total of nine studies which give similar characterizations. Crawford and Bradshaw (Ref. 3, p. 6) reported a study which compared the ranking of 13 cited characteristics as obtained from students, professors and administrators. The study reported that the three groups were quite consistent in their rankings, although there were differences between the groups in the weight attached to each item. The five most important were: (1), knowledge of subject matter; (2), well planned and organized lectures; (3), enthusiasm for teaching; (4) willingness to help students outside of class; and (5), opportunity for class discussion. The four least important characteristics were: (1) use of visual aids; (2), sense of humour; (3), neat appearance; and (4), punctuality. There is some indication that students' perceptions of teaching may be somewhat different in different disciplines; it has been reported (Ref. 3, p. 7) that students of the physical sciences tend to emphasize presentation and organization of subject matter, whereas humanities students emphasized enthusiasm and ability to encourage students.

In the trial period, redundancy can be checked by computing the correlations of the statements taken in pairs. Nondiscriminatory or insensitive items will tend to generate scores for a representative sampleof teachers which are confined to a narrow range of the scale. This in turn makes the results difficult to interpret so that it is important to try to eliminate them. One insensitivity which is particularly difficult to eliminate is the students' tendency to be lenient so that considerably more than 50% of professors receive ratings above the midpoint of the scale. This

decreases the capacity of the rating system to discriminate levels of teaching ability at the excellent end of the scale. Since the basic problem is the provision of adequate recognition of good teaching (Ref. 9, p. 216), it is very important to eliminate items that generate uniformly high scores. Typically, some professionals (Ref. 3, p. 9) recommend that items that focus on the professor's knowledge of his subject not be included in the questions selected for the questionnaire, since they tend to receive a uniformly high rating. Presumably this occurs because nearly all professors appear to be highly knowledgeable from the students' viewpoint. An example of a questionnaire constructed by this process for the specific purpose of rating teaching ability is given in Appendix B. This is used in the Faculty of Social Science at the University of Western Ontario. Note that it calls for responses on nine characteristics together with a tenth question on overall rating.

Other formats can be used for soliciting opinions. For example, one could solicit a response in the form of an essay instead of insisting on a forced choice of answers to specific questions, and some professors believe that teaching is so complex that it is the only appropriate format. Apart from the difficulty in assessing such responses consistently, there is a philosophical problem associated with the open-ended approach. The student is, in effect, selecting his own personal set of criteria or traits and then rating these, albeit usually more crudely than is implied by the set format of a questionnaire.

The administration of the questionnaires should obviously be done in such a way as to ensure that sources of error and bias are minimized. There is general agreement that, when the results from a questionnaire are used for promotion and tenure, considerable care should be exercised in ensuring that procedures are identical in all departments (Ref. 3, p. 13). For example, there is evidence that significant differences can occur between results obtained with the lecturer physically present or absent from the room while the students fill out the questionnaire (Ref. 3, p. 15) and that the results differ somewhat depending on the specific instructions. It has also been suggested that the students tend to be more lenient if they are told that the results are to be used at promotion and tenure proceedings (Ref. 15). Since the distribution of answers is already somewhat skewed to the favourable end of the scale, elimination of inadvertently introduced differences such as those noted above is obviously very important. The scheme usually suggested is to have a third person administer the questionnaire while the lecturer temporarily absents himself from the room. This is done during some specific period, say the last two weeks of the term. The students should be given a standardized set of written instructions, and the administrator should direct attention to the objectives.

It must be appreciated that the typical TEQ is designed to assess performance in the lecture room, although this is frequently not explicitly understood by either professors or students. The extent to which this is a limitation is an important question which will be discussed below. It is to be noted that most professors who excel at teaching rely largely on the lecture, and are likely to continue to do so. In spite of growing interest in other methods, such as modular instruction (Ref. 12), the lecture is likely to continue being used extensively because of the extra costs thought to be associated with these methods.

5. AVAILABLE EVIDENCE ON RELIABILITY AND VALIDITY

We now return to the central questions of the relationship of TEQs to objective measures of learning and to other interested groups' perceptions of teaching. In psychological terms it is necessary to establish both reliability and validity. *Reliability* is defined here as the extent to which a test yields consistent results for a given individual. *Validity* is the extent to which it actually

measures the trait or traits for which it has been designed. The tests for validity can be divided into two classes; direct and indirect methods. Direct methods attempt to compare results of teaching evaluations with measures of learning impact. Indirect methods, which have been widely used owing to the difficulties surrounding the design of an adequate direct experiment, either attempt to determine the extent to which results are affected by irrelevant variables such as class size and severity of marking, or look for correlations with the ratings determined by other suitable groups such as a professor's peers, alumni and paid observers. Seeking correlations with the opinions of other groups also has an importance beyond being simply a substitute for direct methods. It has been noted that important aspects of teaching may not be covered by direct tests. Hence a positive correlation of the results of TEQs and peer opinions would indicate that students are responding to some of the less easily defined qualities that professors feel are important, so that such correlations are extremely important in their own right. Reliability has been investigated by testing for both stability and internal consistency. Stability has most often been checked by correlating ratings obtained a few weeks apart, and such correlations have been found to be quite high (0.80 to 0.90). This suggests that students are not unduly affected by one poor lecture or similar day-to-day variations (Ref. 17). Internal consistency can be checked by such procedures as correlating mean odd-item rating with the mean even-item ratings and, for questionnaires that ask the students to make an overall judgment of teaching effectiveness, correlating the overall evaluation item with the mean of the other more specific items on the questionnaire. Such reported correlations again are always greater than 0.75 and sometimes higher than 0.90 (Refs. 2, 3). Murray (Ref. 15) reported that, for the questionnaire shown in Appendix B, the overall rating correlation was found to be 0.97.

In general most of the tests for various forms of potential bias indicate that they have at most a very small effect. Correlations of the rating given by a particular student with such factors as his grade are usually low. Murray (Ref. 17), in tests on a large multisection introductory psychology course, obtained correlations with assigned mark of 0.20, 0.23 and 0.18 in three successive years. Correlations with a student's academic average seem to show even lower values, typically 0.1 or less. Other bias factors such as rating of the course difficulty, class size, time of class, can show negative correlations but they are nonetheless found to be small, always less than about 0.2 and usually 0.1 or less. Murray concluded that such bias factors might account for only 15 of the deviation in mean teacher ratings. Nichols and Soper (Ref. 18) reached the same conclusions, and Remmers (Ref. 19), in a review of his research on the rating scale used at Purdue since 1927, observed that if 25 or more ratings by individual students are averaged then the ratings have as much reliability as the best available student ability and aptitude tests. Remmers also reports that course difficulty, sex of student or instructor, student grades and personal popularity outside the class do not appreciably affect the rating.

Correlations of student ratings with the ratings and opinions of other groups have usually been found to be at least moderately positive. Costin, Greenough and Menges (Ref. 2) report a weak to moderate relationship with peer ratings (0.30 to 0.65). Drucker and Remmers (Ref. 20) report correlations of 0.40 to 0.68 with opinions of alumni of at least 10 years standing. Murray (Ref. 17), in the study discussed above, reported correlations with peer ratings and those of senior students paid to observe classes of 0.82 and 0.92 respectively. Nadeau (Ref. 13) notes that, although professors and students may have different bases for judgment of good teaching, there is usually considerable agreement as to who is effective. He quotes a study in which both professors and students at one university were asked to nominate the best and the worst teacher they knew. Of the professors who were most frequently cited by both staff and students, there was remarkable

agreement. There was agreement on sixty-six and disagreement on only 10. The breakdown is shown in Table 1 below.

One result which should be noted here is an attempt by Murray (Ref. 17) to correlate an individual teacher's ratings as obtained at different course types and levels. He found that the correlation was rather low (0.52) and observed that a lecturer who rates poorly on an introductory course could fare much better on a senior year honours course. The reverse was also found to be true. The inference that some professors might be most effective for particular course types or levels of student sophistication is hardly surprising. Sheffield notes that, even for a particular course level or type, some professors are more effective with certain types of students and infers that students should be as free as possible in their choice of teachers. In particular he notes that they should not be arbitrarily assigned to course sections (Ref. 9, p. 217).

Category	Faculty's best	Faculty's worst
Students' best	37	8
Students' worst	2	19

Table 1 Student-faculty correlations of teaching effectiveness

Although most of the tests of the type just described indicate that the students rate lecturers maturely, studies have been reported in which different results have been found. Typically, Nadeau (Ref. 13) quotes one study in which it was concluded that high ratings were related to superficial popularity, and cites seven in which no significant correlation was found. And again Costin, Greenough and Menges (Ref. 2) quote one study of ratings obtained in a military academy which reported no correlation between students' opinions and those of the instructors' peers. A recent study by Zeiby (Ref. 21) appeared to show that the ratings can be influenced upward by deliberately teaching in a particular style. He used a "spoon feeding" and "mature" approach in alternate years in two courses, one analytic and one descriptive. The effect was small for some of the questions asked but much larger for others. Typically for a question on preparation for class, on a 1 to 5 scale with 1 most favourable, the rating might improve from 1.67 to 1.3. On another question, which asked if the assignments were well designed, the rating increased from 5.48 to 2.33. The latter variation was obtained as follows. In alternate years in the analytical course, electromagnetism, he adopted two distinct methods of explaining homework assignments. For both cases, written solutions were handed out to the classes. The "spoon feeding" method consisted of explaining these notes on solutions in detail on the board, whereas in the "mature" method he would explore various alternative solutions. The overall rating question was influenced upward by about half a point using the "spoon feeding" approach.

Zelby's work serves to indicate that the questions relating to the reliability and validity of the TEQ are not free of controversy. Indeed, the controversy surrounding results of attempts to obtain correlations with objective measures of learning is particularly marked. Because such approaches are fraught with technical difficulties, they have invariably used standardized tests to compare multiple sections of a large introductory course in which the objectives could be accepted with reasonable unanimity by all the teachers concerned. The reported correlations range from a spectacular *minus* 0.75 (Ref. 22) to as high as + 0.87 (Ref. 23). McKeachie et. al. (Ref. 3, p. 22) quote correlations ranging from – 0.60 to + 0.72 and state an "average" as 0.10. The study by

Murray (Ref. 17) at the University of Western Ontario quotes correlations in the range + 0.21 to + 0.26. As noted earlier, a second major aspect of learning is thought to be the extent to which the lecturer is successful in creating a persisting desire to learn. One way in which the relationship of TEQ results to this objective might be assessed is to ascertain the extent to which students who take an introductory course in a particular subject elect an advanced course in the same subject, and correlate this with their ratings of the lecturer in the introductory course. In other words, is the teacher who receives high TEQ scores more effective in channelling students to his subject area? Only one study along these lines is known to the writer, and it was stated that significant positive correlations were obtained in two out of the five semesters (0.41 and 0.63) in which the study was undertaken, but for the other three, the correlations were effectively zero or negative. The authors noted that the relationship between student ratings and long term student interest "appears to be rather inconsistent" but offered no explanation (Ref. 3, p. 22). In general, it seems that, whereas there is a large measure of agreement amongst professionals on reliability and indirect indications of validity, the subject of direct tests of validity is a matter of heated controversy, so that clearly we can assume that it is not understood. It is perhaps worth mentioning that one theory which has been suggested to explain the negative correlations is that a poor teacher makes the student work harder on his own, while the good teacher lulls the student into a false sense of security (Ref. 15). This sugg- estion is, apparently, not entirely facetious.

In completing this very brief survey, we note some of the areas which are believed to be important and yet for which very little or no work has been done. Obviously much more work on direct experiments is required. There is also no satisfactory data available which establishes whether or not the use of ratings actually improves teaching. It must be noted however that a lack of correlation between the use of TEQs and perceived improvement in teaching could be associated with an inappropriate university reward structure as much as with inadequacies of TEQ procedures. Virtually no work has been done to establish the extent to which students are rating the course while they are supposed to be rating the lecturer. It is noted by Costin, Greenough and Menges (Ref. 2) that existing data on the possible role of "pure entertainment" is inconclusive, although it certainly does not permit a conclusion that entertainment is a major factor. Some attempts have been made to relate ratings to established personality traits, but only very recently were any meaningful results reported. This study, by Murray (Ref. 24), was apparently the first to have the lecturers' personality ratings done by a selection of peers rather than on a self-evaluation basis, so it hardly seems surprising that previous attempts were inconclusive. Murray noted that the four personality traits which had the strongest correlations with overall ratings were leadership (+ 0.57), extroversion (+ 0.52), authoritarianism (− 0.51) and anxiety (− 0.47).

What overall conclusions can one draw? To date, the attempts at direct validation have failed to give convincing evidence of a strong correlation between the results of TEQs and student learning, so one could use this and the other negative evidence cited here to enter a "not proven" judgment. Also, it has been noted by Murray (Ref. 24) that the high correlation of good ratings with a few personality traits might suggest that students are rating the lecturer's personality, and this in turn might mean that his teaching ratings are largely predetermined independently of any effort on the part of the lecturer. But one cannot ignore the large body of indirect evidence; a lecturer who is judged by his peers to be a master of his subject, and by the students to be relatively well prepared, to explain concepts clearly and so on, is obviously making a serious attempt to teach well in the currently accepted sense of the term. And as Murray (Ref. 24) has suggested, an alternative explanation of the high correlation of certain personality traits with TEQ ratings is that these traits

are related to classroom teaching skills which in turn are accurately rated by the students. In general, most of the tests for bias indicate that TEQ procedures are reasonably free of irrelevant factors. Murray (Ref. 25) has noted that Zelby's study is one of the few in this class which has given adverse results, and he has questioned the adequacy of Zelby's procedure on the grounds that Zelby served as both investigator and subject.

An interesting sidelight of Murray's study (Ref. 24) is that it suggests that attempts to train university teachers should place emphasis on developing certain personality traits such as displaying leadership, avoiding excessive authoritarianism and the like. This tends to support a very interesting major conclusion by Sheffield, who stated (Ref. 9, p. 215):

> attitudes towards the students and teaching are more important than methods and techniques.

One likely interpretation of the apparent conflict between the results of the direct and indirect experiments is that student performance in a particular course is the result of a complex interaction with many factors, of which the excellence of the lectures associated with that course is only one (Ref. 10). It is possible that the interaction that actually occurs in the lecture room only very weakly affects academic performance, so that the basic value of the role of the lecture itself can be called into question. For example, it may be possible tor a professor who never gives a lecture to organize his resources in such a way as to produce high levels of attainment and morale (Ref. 12). This in turn suggests that perhaps the current debate on teaching should not be so much concerned with the role of the TEQ as with the role of the lecture in the teaching process.

6. PRACTICES AND ATTITUDES IN THE FACULTY

As noted earlier, TEQs are already widely used in this Faculty. In only one of the nine departments and divisions is there no policy and no known use by individual professors. In one other department their use is voluntary, and apparently not very widespread. In all others, their use is compulsory, but the details of administration vary considerably. In some cases the questionnaires are handed out by the professors, and the department head will intervene and distribute them if the professor refuses to cooperate. In at least two cases, the students administer the questionnaire, and in such cases the distribution and the timing of its filling out can vary widely, and the rate of recovery is often low. In one case brought to the author's attention, the questionnaires for all courses applicable to a given program were given out and filled in together.

The details of the questionnaires vary greatly in style and content. A few have been modelled on the one recommended recently by the Commission on University Government. With a couple of notable exceptions, no clear distinction is made between the sections of the questionnaire relating to teacher, course, and laboratories. No explanation of the intended uses of the questionnaires is made on the documents. In exploring teaching ability some ask only for an overall rating, while others seek more specific responses. In most cases unsigned free comments are solicited, but the way in which they are used varies greatly. The comments are usually passed on to the lecturer, or if he is responsible for collection, he keeps them for his own personal use. In one department, the lecturer is required to pass on a summary, while in two cases the free comments are interpreted by the departmental or division head, and a condensation is filed and is passed on to the lecturer. The results from TEQs, when available, have for several years been regularly tabled at promotion and tenure meetings, and it was stated by several departmental chairmen that they have occasionally proved to be extremely useful in this context.

Amongst the academic staff, attitudes as interpreted from a questionnaire circulated by the

Teaching Methods and Resources Committee seem to be somewhat mixed and perhaps even a little inconsistent. For example about two thirds of the staff appear to feel that the notion that teaching performance can be measured is not meaningful, yet they are also about 2 to 1 in favour of the use of student evaluations for improving teaching. A majority (6 to 4) are concerned that TEQs will promote popularity-seeking habits in the classroom. The faculty appeared to be strongly in favour, by about 6 to 1, of the statement which suggested that the exact form and use of the questionnaire should be discussed and agreed upon amongst the staff before being used. The statements on the questionnaire and the responses are summarised in Appendix C.

Two comments can perhaps be made. Given the uncertainties that still exist, and the insensitivity of the scale at the high end, the range of questionnaire types and administrative practices leaves the unsettling feeling that an unnecessary element of error is being introduced. Secondly, given that a properly designed questionnaire should seek measured responses to attributes which have been carefully selected, the soliciting of unsigned free comments seems somewhat like adding noise back into a process specifically designed to separate out this noise component. Although individual faculty members should no doubt be free to seek such comments and interpret them as they please, their use in any official capacity does not appear to make much sense. In fact there is some evidence that the comments tend to have a distribution which is skewed to a negative attitude, so that the effect in terms of staff morale can be out of all proportion to any meaningful gain in information.

7. CONCLUDING REMARKS: SUGGESTIONS FOR POLICY

In general, the difficulties still surrounding the validity of the TEQ are such that there is room for a range of opinions as to their role in the recognition of, and improvement in, teaching quality. This writer feels that, on balance, there is sufficient positive evidence to warrant their use in a suitably cautious fashion. The situation appears to be that we do not yet have anything better available; evaluations by department heads seem particularly prone to error. Their opinions are often based on colleague opinion, and as Sheffield has stated,(Ref. 9, p. 217)

> professors observe the equivalent of film clips of each other teaching, while students take part, as it were, in feature length movies.

It is hard to avoid the conclusion that a systematic assessment of student opinion is a useful step forward.

There are, however, two very obvious dangers associated with their use which should be reflected in any policy adopted. The first relates to the inevitable tendency to attach undue significance to the numbers involved. It is clear from the data quoted above that there are sources of error which could easily shift the rating by at least 1/2 point and possibly much more. Murray's study (Ref. 17), which showed the rather poor correlation between ratings given by first and senior vear classes, is a typical indication of the serious limitations of TEQs. A simple averaging of TEQ results obtained in a variety of courses may well be unjust. Also Sheffield has noted that teacher and course evaluation procedures ought to be sophisticated enough to take into account the fact that professors will tend to be more effective for certain student types. In spite of the usual protestations by Faculty administrators, it seems that such over-interpretation will inevitably occur. Indeed there is evidence that in one case (not in this Faculty) ratings are being used directly in merit increase formulae. Such an approach is clearly an abuse.

The second danger, as almost all of the writers in the field strongly emphasize, is that student

ratings of instruction fall far short of a complete assessment of a professor's teaching contribution. Some of the empirical data itself shows that the typical undergraduate student has a limited perspective of a professor's fitness to teach. The student's tendency to rate professors uniformly highly on their knowledge of subject matter is a clear indication that they are not able to make a mature judgment as to the appropriateness of the material selected for a course. They also apparently consider active research or professional involvement as being largely irrelevant, as is shown by the rank ordering of traits reported by Perry (Ref. 10). Item #8, "publishing material related to his subject field" was ranked sixtieth at two universities. Yet almost all would agree with Sheffield's observation that active involvement in the development of his field is an essential attribute of the truly outstanding teacher. Costin, Greenough and Menges (Ref. 2) conclude that,

> Other obvious factors which should be taken into account in any overall measure of instruction include participation in thesis committees and direction of graduate research, teaching awards given by students, individual undergraduate instruction and research direction, department colloquia, participation as a guest lecturer in other courses, and the development of new courses or improving the materials and methods in existing courses.

The apparently very weak correlation of learning with a teacher's performance in the lecture room serves as very clear warning that these other factors are obviously very important, possibly even more than lecturing itself. Any evaluation system which does not give, or is not seen to give, these other contributions their due weight is obviously unjust.

It is appropriate to make some suggestions relating to development of a policy for this Faculty. Because of uncertainties still surrounding TEQs and the established tendency towards bias in certain aspects, the variety of designs and administrative procedures now in use cannot be considered acceptable for promotion and tenure purposes. An approach which would help eliminate many of the problems cited here, and yet which would be flexible, is to have two questionnaires; one administered at the Faculty level for promotion and tenure considerations and one at the departmental level with a wider range of objectives. The Faculty level questionnaire would be shorter and procedures could be kept uniform for all classes. The departmental level questionnaire can be designed and administered by the department to suit its particular needs. For example, in addition to seeking comments on course content, departmental questionnaires might be designed to provide much more diagnostic information to the individual teacher than is possible with a short questionnaire used by the Faculty administration.

Although there is a fairly large amount of evidence in favour of TEQs it seems to this writer that, any official policy notwithstanding, the level of controversy is such that the use of TEQs should not be forced on the small minority in the Faculty who object strongly to them. A suitable procedure could be devised to ensure that all Faculty professors are automatically included with the exception of those who specifically request to opt out. It would seem that such an approach is entirely reasonable. Presumably since the majority are persuaded that TEQs have some value, then at promotion and tenure proceedings, absence of TEQ records would be the equivalent to having a weak publication record. The professor has to make his case on a more restricted data base.

By far the most important, there should be a clearly developed Faculty policy which recognizes the other very real facets of teaching not covered by TEQs. There is the very real risk that, now TEQs are gaining acceptance, we will move from a position of paying lip service to teaching, to paying lip service to the task of seriously evaluating all its aspects. This policy must make every effort to counter the undue significance that will inevitably attach to the numbers produced by TEQ

procedures and the obvious tendency to avoid attempting serious in-depth evaluations. It has been noted (Ref. 5) that we as professors do not shrink from making very thorough evaluations of a man's research contribution. Nor, it is hoped, are we naive enough to simply count numbers of publications without reference to the quality of the journal and outside opinions as the quality of the papers. In evaluating a man's research we make essentially subjective judgments based on as many indicators as possible. The implication is clear that equivalent procedures for evaluation of teaching should be developed. The memorandum issued to the academic staff in November, 1973, described the attributes associated with an outstanding teaching rating in the following terms:

> Knows subject thoroughly; well prepared and well delivered courses; inspires students; good judgment in examining and evaluating; generous with help and counselling; studies subject to keep up to date; is innovative in teaching methods; writes text books; can teach wide range of subjects. (Ref. 26)

This is a useful first step. However much more attention has to be paid to the development of an appropriate data base.

In conclusion, the author would like to return to a theme raised earlier. His personal opinion is that much more effort is required in understanding the role of the lecture in the teaching process. Traditionally the lecture has been basically a method of communicating information, as it had to be in less affluent societies. Now, with the widespread availability of relatively cheap text books and methods of reproduction of lecture notes, the usefulness of this aspect of lecturing can be called into question. But discussion of this problem is beyond the scope of this essay.

ACKNOWLEDGEMENTS

Completion of this essay would have been much more difficult were it not for the advice, assistance, and active encouragement of a number of people. Professor I. R. Dalton undertook most of the work involved in the preparation and interpretation of the questionnaire. Professors I. I. Glass, P. C. Hughes and P. C. Stangeby gave many useful comments. The technical comments given by Professor H. G. Murray of the Department of Psychology, University of Western Ontario were very much appreciated.

REFERENCES

1 — , University of Toronto: Policy and Procedures on Academic Appointments.
2. Costin, F., Greenough, W. T., and Menges, R. J. 1971. Student Ratings of College Teaching: Reliability, Validity and Usefulness. *Review of Educational Research*, **41**, 511-535.
3. Murray, H. G. 1973. *A Guide to Teaching Evaluation*. Ontario Confederation of University Faculty Associations, April.
4. Andreski, S. 1972. *Social Sciences as Sorcery.* London: Andre Deutsch.
5. Good, H. 1975. Ontario Universities Program for Instructional Development. Newsletter No. 4, January.
6. Hexter, J. H. 1969. Publish or Perish—A Defence. *The Public Interest*, **17**, 60-77.
7. Hayes, J. R. 1971. Research, Teaching and Faculty Fate. *Science*, **172**, 227-230.
8. McDaniel, E. D., and Feldhusen, J. F. 1970. Relationships Between Faculty Ratings and Indexes of Service and Scholarship. *Proc. 78th Annual Convent ion of the American Psychological Association*, 5, 619-620.
9. Sheffield, E. F. (Editor). 1975. *Teaching in the Universities - No One Way*. Montreal and

Kingston: McGill-Queen's University Press.
10. Falk, B., and Dow, K. L. 1971. *The Assessment of University Teaching*. London: Society for Research into Higher Education.
11. Perry, R. R. 1969. Evaluation of Teaching Behaviour Seeks to Measure Effectiveness. *College and University Business*, **47**, 18-22.
12. Kulik, J. A., Kulik, C. L., and Carmichael, K. 1974. The Keller Plan in Science Teaching. *Science*, **183**, 379-383.
13. Nadeau, G. G. 1974. Research on Students' Evaluation of College and University Teaching: The Weight of Evidence—Updated. Seminar on Evaluation, Faculty of Education, University of of Toronto, February.
14. Crawford, P. L. 1968. Perceptions of Characteristics of Effective University Teachers: A Scaling Analysis. *Educational and Psychological Measurement*, **28**, 1079-1085.
15. Murray, H. G. Private Communication.
16. Furdy, C. Private Communication.
17. Murray, H.G. 1973. The Reliability and Validity of Student Ratings of Faculty Teaching Ability. *Journal of Educational Measurement*, January.
18. Nichols, A., and Soper, J. C. 1972. Economic Man in the Classroom. *Journal of Political Economy* **80**, 1069-1073.
19. Rermners, H. H. 1959. On Students' Perceptions of Teachers' Effectiveness. *The Appraisal of Teaching in Large Universities*, W. J. McKeachie (Editor) University of Michigan.
20. Drucker, A. J., and Reminers, H. H. 1951. Do Alumni and Students Differ in Their Attitudes Towards Instructors? *Journal of Educational Psychology*, **42**, 129-143.
21. Zelby, L. W. 1974. Student-Faculty Evaluation. *Science*, **183**, 1267-1270.
22. Rodin, M., and Rodin, B. 1972. Student Evaluations of Teachers. *Science*, **177**, 1164-1166.
23. Frey, P. W. 1975. Student Ratings of Teaching: Validity of Several Rating Factors. *Science*, **182**, 83-85.
24. Murray, H. G. 1973. Predicting Student Ratings of College Teaching From Peer Ratings of Personality Traits. Presented at Midwestern Psychological Association Conference, Chicago, May.
25. Murray, H. G. Comments in a review of a draft version of this essay.
26. Etkin, B. 1973. Memorandum to Academic Staff Discussing Dimensions of Assessment. Faculty of Applied Science and Engineering, Nov. 20.

APPENDIX A

Rankings of 60 Possible Criteria of Good Teaching by Students at Four Universities

(Source: Ref. 11)

NMS: New Mexico State, Las Cruces.
NIU: Northern Illinois University.
WKU: Western Kentucky University.
UT: University of Toledo.

The number of responses on which the survey is based:
NMS = 654; NIU = 2488; WKU = 1698; UT = 1793.

RANKINGS

	Criterion Behaviour	NMS	NIU	WKU	UT
1.	Evidencing better than average speech qualities	26	25	27	26
2.	Constructing tests which search for understanding on the part of the students rather than rote memory ability	4	5	9	5
3.	Providing several test opportunities for students	27	28	29	32
4.	Engaging in continued formal study in his field	24	29	31	28
5.	Acknowledging all questions to the best of his ability	12	12	14	12
6.	Motivating students to do their best	11	9	5	10
7.	Explaining grading standards	40	37	42	45
8.	Publishing material related to his subject field	57	59	60	60
9.	Having practical experience in his field	20	19	24	21
10.	Communicating effectively at level appropriate to the preparedness of students	7	6	6	7
11.	Identifying his comments which are personal opinion	28	27	41	27
12.	Challenging students' convictions	44	38	52	43
13.	Utilizing visual aids to assist in creating subject matter achievement with students	47	48	45	47
14.	Announcing tests and quizzes in advance	39	41	36	46
15.	Making written comments on corrected returned assignments	22.5	17	26	25
16.	Presenting organized supplementary course material	43	42	47	41
17.	Establishing good rapport with students in classroom	17	15	15	17
18.	Making an effort to know students as individuals	36	30	28	38
19.	Inspiring students to continue for graduate study	52	52	51	49
20.	Demonstrating comprehensive knowledge of his subject	6	10	10	3
21.	Exhibiting an intelligent personal philosophy of life	46	44	38	40
22.	Encouraging student participation in class	25	22	23	24
23.	Beginning and ending classes on time	48	51	48	51
24.	Accepting justified constructive criticism by qualified persons .	22.5	21	21	23

25.	Sharing departmental duties with his colleagues	50	49	49	50
26.	Having irritating personal mannerisms	53	54	57	54
27.	Establishing sincere interest in subject being taught	2	2	3	2
28.	Taking measures to prevent cheating by students	38	43	32	31
29.	Recognizing his responsibility for the academic success of students	21	26	17	18
30.	Devoting time to student activities on campus	59	58	54	58
31.	Demonstrating a stable level-headed personality	35	31	22	30
32.	Returning graded assignments promptly	30	32	34	34
33.	Patiently assisting students with their problems	16	18	13	20
34.	Holding membership in scholarly organizations	55	55	56	56
35	Being well prepared for class	1	1	1	1
36	Setting high standards of achievement for students	18	23	25	16
37	Involving himself in appropriate university committees	58	56	55	55
38	Being knowledgeable about the community in which he lives	54	53	53	53
39	Being readily available for consultation with students	14	14	15	16
40	Displaying broad intellectual interests	41	36	40	36
41	Treating students with respect	10	4	2	11
42	Raising the aspirational level of students	19	20	18	17
43	Being able to show practical applications of subject	13	13	12	13
44	Organizing the course in logical fashion	8	11	11	9
45	Making appearances which assist programs of community	60	60	58	59
46	Earning the respect of his colleagues	45	45	37	42
47	Encouraging intelligent independent thought by students	5	7	8	8
48	Using teaching method's which enable students to achieve objectives	9	8	7	4
49	Rewriting and updating tests	15	16	19	14
50	Presenting an extensive lucid syllabus of the course	49	46	50	48
51	Explaining grading procedures	37	34	39	41
52	Being consistently involved in research projects	56	57	59	57
53	Seldom using sarcasm with students	34	47	43.5	39
54	Indicating that the scope and demands of each assignment have been considered carefully	33	35	35	33
55	Being fair and reasonable in evaluation procedures	3	3	4	6
56	Relating course material to that of other courses	31	40	46	35
57	Using more than one type of evaluation device	29	24	30	29
58	Being neatly dressed	5l	50	43.5	52
59	Exhibiting a genuine sense of humor	42	33	33	37
60	Encouraging moral responsibility in students by example	39	20	22	32

APPENDIX B
Questionnaire used in the Faculty of Social Science at the University of Western Ontario

GENERAL INFORMATION GIVEN TO STUDENTS

In this Questionnaire you are asked to rate your instructor's abilities as a teacher. A statistical summary of your ratings, combined with your classmates' ratings, will be used both for feedback to the instructor and as one of several sources of information to be considered in decisions relating to faculty tenure and promotion. Therefore, it is important that you make your ratings in a completely honest and responsible way. To ensure confidentiality, your instructor should not be present while students complete the Questionnaire, and he should not be involved in the distribution and collection of Questionnaire materials. Also, students are asked not to identify themselves by name or student number or other means on their response cards. Questionnaire responses will be processed by the Social Science Computing Laboratory. Statistical summaries will not be released to instructors prior to the first week of June. Individual student ratings will never be released. Thus your responses cannot and will not be used to affect your standing or progression in this or any other course. Once released, statistical summaries will be made available to the University Students' Council.

QUESTIONS AND RATING SCALES

(Rating scale - Questions 1 to 9: 5 - strongly agree
4 - agree
3 - undecided
2 - disagree
1 - strongly disagree)

1. The instructor is a good speaker.
2. The instructor is well prepared for classes.
3. The instructor presents his material in a well-organized and coherent manner.
4. The instructor is able to explain difficult concepts a clear and straightforward way.
5 The instructor makes effective use of examples and illustrations in explaining course materials.
6. Considering limitations due to class size, the instructor does a good job of answering questions that are asked in class sessions.
7. The instructor is enthusiastic about his subject matter.
8. Considering inherent limitations of the course content the instructor is successful m presenting the subject matter in an interesting way
9. The instructor is successful in encouraging students to think independently and do supplementary reading related to the subject matter of the course

(Rating scale - Question 10: 5 - outstanding
4 - very good
3 - good
2 - satisfactory
1 - poor)

10. How would you rate your instructor in terms of general, overall effectiveness as a teacher?

APPENDIX C
Survey of Academic Staff Opinion

A brief sounding of the opinion of the Faculty professorial staff was undertaken in order to obtain some idea of attitudes, so that the style of the present document would be appropriate to these attitudes. It was not intended to provide a definitive basis for policy decisions. The survey solicited responses to twelve statements which were believed to be representative of the ranges of opinion encountered in this subject. The questions together with the responses are given below.

QUESTIONS

1. Teaching is more art than science and the notion that meaningful parameters of good teaching can be identified and measured is a gross oversimplification.
2. The submission of free comments bv students, after the answering of a well-designed set of questions, merely introduces noise into an otherwise well-behaved evaluation process and should be dispensed with.
3. Evaluations by students should be encouraged as a means of improving teaching quality.
4. The soliciting of free comment outside the format of the set questions on a student questionnaire allows the student? an opportunity to identify problem areas that are not covered by the standard questionnaire, and should be encouraged.
5. The publishing of teaching evaluations in any form is a potentially libellous practice and should not be allowed without the specific permission of the staff member involved in each course.
6. Student evaluation of teachers and courses provides the teacher and the curriculum committee with some very useful feedback, and therefore a regular and uniform evaluation of all courses should be conducted by the faculty office.
7. The use of the results of classroom evaluation of teaching for promotion and tenure considerations is a significant step forward in achieving adequate recognition and reward of teaching.
8. Teaching evaluations can measure only classroom performance, and their use will promote popularity-seeking, not better teaching.
9. It is preferable that students be asked to describe a lecturer's performance rather than be asked to answer a questionnaire.
10. If the student evaluation of teaching is to be adopted and used in deciding questions of promotion and tenure, the exact form of the questionnaire and the specific ways in which the results are to be used must be discussed and agreed upon by at least two-thirds of the academic staff of the faculty.
11. If a lecturer wants to know how his class feels about his course or his handling of it, then he should be free to ask, in whatever way he wishes, and the results should be for his use only.
12. Allowing the students these opportunities to share in curriculum plan- ning and staff promotions will promote good morale among our students.

RESPONSES BY PERCENTAGE

Number of Respondents = 91

Question	Strongly agree	Agree somewhat	No opinion	Disagree somewhat	Strongly disagree
1	11	43	7	23	16
2	8	15	6	33	38
3	58	27	2	8	5
4	45	37	6	15	8
5	41	30	6	15	8
6	32	38	8	10	12
7	24	42	3	13	17
8	12	44	4	32	8
9	5	16	20	29	30
10	41	27	19	9	4
11	35	24	25	11	5
12	19	29	31	9	12