

Interest in programs which purport to yield judgements about the worth of a professor *qua* teacher from statistical analysis of student opinion is fairly widespread. Interest often goes as far as the explicit use of such data as the major basis upon which to distribute tenure and promotion and other elements of the academic reward system. While the attitudes of students, faculty, and administrators are often comprehensible, they do tend to lead to contradictory conclusions about the worth of such programs because each fails to see the position of the other and the incompleteness of their own position.

Nathan Kroman*

Student Evaluation of Professors: Critique and Proposal

Introduction

Interest in programs which purport to yield judgements about the worth of a professor *qua* teacher from statistical analysis of student opinion is fairly widespread. Interest often goes as far as the explicit use of such data as the major basis upon which to distribute tenure and promotion and other elements of the academic reward system. While the attitudes of students, faculty, and administrators are often comprehensible, they do tend to lead to contradictory conclusions about the worth of such programs because each fails to see the position of the other and the incompleteness of their own position.

This paper reflects the author's undefended prejudice that the ideal of a university as a citadel of intellectual power and integrity should be operative in university affairs as well as in laboratories and classrooms. Intellectual power includes the ability to differentiate between what is known and what is not known. Intellectual integrity implies the avoidance of policy based on whim or expediency. I would like to offer a fairly accurate picture of the motivations and reasonings of the three groups regarding class evaluation followed by a critique of all three positions. Following this will be a brief conception of an alternate teacher evaluation system in which the legitimate positions of each can be served in a way that has more intellectual power and integrity than most of the current systems. It should be made clear that a teacher evaluation system includes not only data gathering, data processing, but data interpretation, and data usage. Although what is being proposed is different than many existing systems, its chief uniqueness is in the area of data interpretation and data usage.

The thinking in this paper was provoked by the assignment to develop a teacher evaluation program at a Canadian university. Involvement included lengthy discussions with students, faculty, and a host of deans and department heads. It also led to the examination of about fifty teacher evaluation programs which were in use at varied North American institutions. Among other preliminary conclusions were the following: (1) Large numbers of programs failed to receive agreement among those affected regarding the purposes or the procedures. Compromises led to rationales and purposes which were logically incoherent. (2) There was a lack of coherence between stated purposes and procedures which include not only instrumentation and statistical processing, but include the use to which data was put. (3)

*Nathan Kroman, Ph.D., Wichita State University

There was a remarkable lack of recognition about what is reliably known about teaching and even less respect for what is not known about teaching. (4) Embedded in many programs and the controversy that surrounds them were some significant but unexamined normative and descriptive assumptions about the nature of education itself.

Differential Attitudes

Students have provided much of the impetus for teacher evaluation schemes at many institutions. In these cases, students felt dissatisfied with classes and instructors and also felt that administration and faculty were insensitive to their concerns. Student groups have developed their own procedures and produced "consumer reports" for the benefit of new students as a weapon against "bad" teachers.

Administrators tend to support these programs as a highly significant basis upon which to identify and reward "effective teaching" in response to pressure from boards, political groups, and students. Support, *I know*, is sometimes *privately* defended within administration conference rooms as a device to placate students who see themselves as customers rather than clients in this era of declining enrollments.

Faculty, on the other hand, tend to oppose these procedures on the broad ground of professionalism. This includes such views as (1) students are generally not qualified to judge the expertise of a professor, (2) tying the university reward system to student approval is pressure towards intellectual mediocrity, (3) specific instruments and statistical analysis do not meet acceptable standards of validity and reliability, (4) class evaluation is customer rather than client oriented, and (5) class evaluation violates academic freedom or autonomy since it places external limits on a professor's ability to conduct his class as he feels it is appropriate.

Critique

To argue that students are unaffected by their classes and instructors would be to argue for the abandonment of universities and colleges. Justification of the claim that students should have input into systems that affect them seems unnecessary in this age of women's lib, the minority movement, and third world politics. It is, however, necessary to remember that students, like the rest of us, do not always know what is in their own long term interests. We all tend, from time to time, to confuse pleasantness with meaningfulness.

Fuel to the students' fire is unfortunately provided by their recognition of what most of us have always known but seldom acknowledge except to each other. Not all professors are concerned with education *per se*. Some merely accept teaching as the price they are willing to pay in order to be in an environment in which they can practice their disciplines. The facetious quip "We mustn't let students interfere with the real work of the university" heard at a prestigious university was clearly not always facetious. Similarly, all too many students see university administrators as primarily concerned with the maintenance of peace and a smoothly operating *status quo* rather than with genuine educational growth and inquiry.

The pressure on administrators is real whether coming from the current and well ridden hobby horse of Skinnerian accountability or from the more traditional limitations of finite resources. To administer and accept responsibility for administration involves the practical but highly uncomfortable necessity to make differential judgements of competence. The relatively smooth road to the

apparently firm ground of statistics, numbers, and measurement and their aura of "science" provides an illusion of objectivity that is understandably attractive. Not the least of its attractiveness is to be able to cite student opinion and thereby pass the responsibility for uncomfortable decisions to the customer. This may be warranted if we are selling shaving cream.

Although many teachers, professors, and administrators speak glibly of "effective teaching" there is precious little agreement on the philosophical or operational definition of the operant adjective. Moreover, there is less than perfect agreement on the definition of the noun. If "effective" is used in its ordinary language sense referring to the quality of achieving a desired objective as a consequence of the planning and procedures of the teacher, the usual teacher evaluation techniques do not and can not measure the effectiveness of a teacher.

Even if "effectiveness" were measurable at the conclusion of a class by a student questionnaire, few of us, upon careful and critical reflection, would judge a teacher worthy of reward if he were merely effective. The value of effectiveness can never be greater than the value of the objectives. Professors, within the broad and varied limits of their respective disciplines, their own mastery of that discipline, within the limits of the catalog description of their courses, and within the extremely broad limits of institutional objectives do have the legal and practical right to define objectives.

Faculty arguments against student evaluation of courses also have the appearance of validity but fail to make a complete case. Students are probably not competent to judge the expertise of a professor in substantive matters relating to his discipline. There is also reason to question whether students are even in a position to judge the long term meaningfulness and significance of their classes at the conclusion of a term. The typical student at that time is probably rather tense, busy, and task oriented towards the imminent *rite de passage*. Most of us know of a few former teachers whose high value was appreciated long after we left them.

On the other hand, while professional concern for meticulous instrumentation and statistical analysis is ordinarily praiseworthy as an attempt to increase rigor in research, it loses much of its credibility when directed against teacher evaluation by faculty members with no more expertise in research related skills than is expected of a concert violinist. The credibility of the attacks on measurement and statistics is further reduced in the light of the procedures which large numbers of professors use in grading students. This process is often as consequential for the student as class evaluation is for the professor.

Professors are, of course, considered professionals but to claim the same autonomy that the fee taking professionals such as law or medicine now enjoy seems unwarranted. Even if professors had the autonomy of the fee takers, it is hard to see how an examination of student reactions would, *in itself*, violate that autonomy. On the contrary, some relatively systematic picture of the effect that one's teaching has on students is potentially helpful in the intelligent exercise of that autonomy even as a physician uses procedures to secure feedback from his patients.

An Alternative Strategy

It seems reasonably clear that the previous analysis identifies a dilemma. Students should have an opportunity to provide inputs into systems that affect them yet are not competent to judge the expertise of a professor. If academic programs

and degrees are to be more than a collection of unrelated and individually sanctioned experiences guided by a random assortment of completely autonomous professors then administrators should have the practical responsibility and power to influence the processes of reward distribution. Professors should have some autonomy about what they do if intellectual growth and innovation is not to be stifled. Teaching is not a craft with clear mutually agreed upon objectives nor with well documented empirical proof of the means to achieve them. The distribution of rewards based on the extent to which a professor fits students' current stereotypes about the "good teacher" is more consistent with the ideals of the marketplace than with those of a university.

The concept of "quality" education is defined in so many ways as to preclude any absolute definition. The various concepts are themselves grounded, more or less rationally, on a complexity of latent and manifest assumptions which are academically classified as Sociological, Political, Philosophical, Economic, Psychological, and many others. If we are to institutionally define "quality teaching" for the purposes of reward distribution, the previous analysis suggests that we may have to choose defining it as (1) what the dean or department chairperson says it is, (2) what the students say it is, or (3) permit intelligent diversity among faculty who accept the responsibility to defend their conception in the light of philosophical and empirical considerations which are themselves respectable.

Each professor should minimally be expected to have explicit intentions with regard to the outcomes of his teaching. Failure to develop concern for teaching to that point is probably the only teacher characteristic that most of us would accept as indicative of genuinely poor teaching. Having assumed that a professor has clear and explicit intentions with regard to the outcomes of his teaching, we can also assume that he has a self-image with respect to what he is doing in class as well as some conception of the image his students have of him. He probably also has a latent, and therefore unexamined, notion of the relationship between these images and his objectives. He probably would accept the notion that the way in which his students see him has some influence on his effectiveness. He would also agree that serious discontinuities between how they see him, how he thinks they see him, and how he'd like to be seen would seriously impair his effectiveness. It is therefore hypothesized that some reasonably reliable indication of how he is seen by his students with regard to a number of carefully selected and discrete behavior dimensions can be useful in improving the quality of his teaching as *he currently defines quality*.

Instruments which tap students' perceptions should thus be worded in descriptive terms although the line between the descriptive and the evaluative is admittedly thin. The wordings should attempt, as much as possible, to yield data which describes the student's perception of what a professor does rather than tap students' judgements of the worth of what he does. The major point is that the data is to be interpreted as students' perceptions and not as indices of worth. No attempt to collapse data into any overall index such as in factor analysis should be undertaken.

One system under development has the potential of identifying what we may call a Population and a Subject. Populations may be defined as an entire university, a college, a department, a course, or the pattern of responses of a particular professor over time. A Subject may be defined as all of the foregoing except the entire university and with the addition of a particular class section of a professor at a given point in time. Data is stored cumulatively. Output provides two lines per

item. The first is the data for the designated Population group. The second is the data for Subject group. For each we read the percentage of students who chose each of the alternatives provided by an item and the mean and standard deviation of the responses. In addition, a Goodness of Fit Chi Square is computed and printed testing the probability that the Subject group pattern deviates from the Population group by chance. The system also identifies characteristics of each student which offers the possibility of defining discrete student characteristics as either Population or Subject groups.

The first two to four occasions on which a professor uses such an instrument, the output separately defines each of his sections or classes as Subject and the entire data bank as Population. The data is kept completely confidential to himself. He then has the potential to make whatever changes in his subsequent classes he feels is necessary to secure response patterns that he would interpret as indicative of quality work. It also provides him with grounds upon which to modify or defend his intentions in the light of the limitations discussed earlier.

Evaluation of a professor should be undertaken by those with formal responsibility to do so. They can and should use such descriptive data but only in the context of a dialogue with the professor concerned. The data should then be jointly interpreted in the light of the professor's intentions if indeed he has adequately defended them. Part of the interpretation of data would hopefully include regard for the professor's reasons for his satisfaction of the data in the light of his intentions. Let us examine some perspectives which a professor might bring to such a situation in defense of "low" student perceptions of some rather common dimensions of teacher evaluation instruments. (1) For *some* courses and for some objectives "low" student perceptions of a professor's acceptance or toleration of student opinion may be quite appropriate and indicative of solid teaching. (2) Many instruments ask students to rate the extent to which a class was "well organized." Students may not perceive a class as having a high degree of organization easily apparent to students but which fits the purposes, objectives of the class and which perhaps is an organization better suited to how people learn. (3) Students may correctly state that they have had "little opportunity" to ask questions in class because the professor prefers to handle these in his office on a one to one basis. (4) Students may find note taking difficult because of the professor's teaching style. There are some rather pointed questions that could be asked regarding the worth of note taking particularly when the notes to be taken are material that the professor is reading directly from his notes. There is reason to believe that note taking is counter productive particularly when students feel that they "have the material" when it is in their notes. (5) Students may praise examinations because the questions are "very unambiguous." Such questions merely test rote memory or are written in such a way that they are easily "psyched out." For some courses and content "ambiguous" questions may be the proper ones.

Conclusion

Although we recognize that in the final analysis we cannot completely separate the evaluative and descriptive components in students' perceptions, we can separate them to the point where the distinction can be useful. This, as has been suggested, can be accomplished by the careful wording of an instrument. The distinction can be made even more useful in the processes of interpretation of student perceptions. Students may, in spite of the care with which we word our instrument,

respond in terms of their evaluation. We interpret the data as merely their perceptions and do not confuse their judgement that something is better or worse with it, in fact, being better or worse.

The imperfections in what is being proposed are obvious but there are intellectual and moral grounds for choosing these kinds of imperfections over the imperfections that seem to be operative in other systems of teacher evaluation.