

A WORM'S-EYE VIEW OF THE TEACHER¹

H. H. Remmers
Professor Emeritus, Purdue University

The Rationale of Student Ratings of Teachers

No teacher has any option as to whether he wants to be rated by his students. Such ratings go on continuously, though informally and unsystematically, among his students. The only choice open to him is whether he wants to have these ratings in systematic form and profit from such feedback for his own self-improvement or to provide it as a partial basis for administrative decision-making concerning teacher assignments, salaries, and promotions.

In my philosophy administrative use of student ratings should be only at the option and with full consent of the teacher, on the premise that if imposed by administrators it could seriously damage the intangible thing that we call morale. Such morale if once damaged might be very difficult if not impossible to restore.

And yet the administrator does have to make decisions of the kind I've just alluded to. It's a plausible hypothesis that at least sometimes the gripings of a few disgruntled students may receive undue weight in such decisions. Unfortunately volunteering to be rated is somewhat analogous to temperance lectures and religious revival meetings - those who most need it don't go.

Arguments against student ratings of teachers include the following:

1. Pupils are incompetent to judge the merit of either the process or the results of teaching, incapable of distinguishing between bad and good teaching, and prone to judge what the teacher does rather than what he gets the pupil to do. This argument may be answered on the grounds that even if, as is doubtful, it states the truth, it is important to ascertain pupils' attitudes toward their teachers because they exist and exert a powerful influence on the effectiveness of instruction. The adage, "You can lead a horse to water but you can't make him drink," applies here.

2. Attaching importance to pupil ratings commits the democratic fallacy of implying that teaching is best which pleases the majority of pupils, and that teaching should be

¹Paper given at FERA's Tenth Annual Testing Conference, Clearwater, Florida.

adjusted to achieve this end. This argument may be answered on the grounds that the best educational process is in essence democratic, and the use of pupil opinion makes possible a wholesome kind of cooperative effort to improve the learning situation.

3. Pupils are inclined to make snap judgments that are consequently unreliable. But the available statistical evidence indicates that the average ratings of teachers by a group of pupils about equal in number to those in the average classroom have a reliability as great as or greater than that of most standardized achievement tests.

4. Pupil's judgments of teachers may be affected and distorted by such irrelevant factors as grades, amount of work required by the teacher, the pupil's interest in the subject, the difficulty of the subject, the preestablished reputation of the teacher, the general attitude toward school, and a lack of seriousness in making the ratings. It can be answered that correlational studies have shown little relationship between most of these factors and ratings of teachers; in particular, pupils' grades, attitudes toward subjects, amount of work required by teachers, and general attitude toward school have been found to correlate to only a low degree, or not at all, with their ratings of teachers. It is more difficult to ascertain the effect of preestablished reputation, but insofar as such an effect exists and influences present ratings, it also constitutes desirable evidence concerning a teacher. The lack of seriousness in making ratings would tend to lower the reliability of the ratings; however, since ratings have been found to be reliable, it follows that pupils have in most investigations taken a serious attitude toward this assignment. In any case, it is possible to eliminate this factor by taking effective steps to establish rapport with pupils when the assignment is explained to them.

5. Pupil ratings tend to disrupt the morale of the teaching staff by arousing hostility, self-consciousness, discouragement, disrespect between colleagues, and attempts to cater to adverse pupil opinion through activities irrelevant to good teaching. Whenever such a danger seems to be present, teachers should be permitted to keep their ratings strictly confidential rather than being required to submit them to their administrative officers. This situation did not seem to be present in most of the published reports dealing with the problem.

6. Pupil rating seems to have a disruptive effect on the morale of pupils; they may come to feel that they are the judges of the worth of teachers, curricula, and school activity. No evidence of this has been found in any of the rating schemes whose results have been published. Bowman

(1934) states, on the basis of several years of experience with having student teachers rated by their pupils, that pupil morale is improved by the opportunity.

Arguments in favor of pupil ratings of teachers that provide their rationale and justification are as follows:

1. Attitudes of students toward their teacher is of major importance in the teaching-learning situation.

2. "Since (such scales are) designed to measure the attitudes of pupils, to the extent that reliable measures are obtained they are also valid since we are concerned not with the characteristics teachers actually possess in the sight of some omniscient judge, but with the characteristics they possess in the eyes of the children they teach. In the words of T. L. Kelley (1926, p. 9), "If competent judges appraise Individual A as much better than Individual B as Individual B is better than Individual C, then it is so, as there is no higher authority to appeal to." (Remmers & Gage, 1955, p. 498).

3. "The logic underlying its construction is another argument for the validity of the scale. Insofar as verbalized opinions are measures of attitudes and the scale measures verbalized opinions, it must also measure attitudes." (Remmers & Gage, 1955, p. 498).

4. The cost in time and money of obtaining pupil ratings is low-----generally less than that of administering a standardized achievement test.

Limitations of time prohibit any attempt to summarize the very extensive technical literature on the subject of student ratings of teachers. Those seriously interested will find the following references useful sources for exploring the literature.

Gage, N. L. (Editor) Handbook of Research on Teaching. Chicago: Rand McNally & Company, 1963.

Harris, C. W. (Editor) Encyclopedia of Educational Research. New York: Macmillan Co. 1960.

Ratings at the Elementary Level

In 1942 Sister Mary T. Amatora, O.S.F. completed her doctoral dissertation research at Purdue University in the construction, experimental tryout and validation of two equivalent forms of the Diagnostic Teacher-Rating Scale with myself as her Major Professor. The scale was developed by

asking pupils to list all the qualities they liked and disliked in teachers. From several thousand items thus obtained seven broad categories emerged:

1. Liking the teacher
2. Ability to explain
3. Kindness, friendliness, and understanding
4. Fairness in grading
5. Discipline (keeping order with the children)
6. Amount of work required
7. Liking for lessons

Under each of these headings are listed seven qualities ranging from very desirable to very undesirable and experimentally scaled with approximately equal intervals.

Reliability of the scale has been found by Tschechtelin, Hipkind and Remmers (1940) using the split-test procedure to range from .86 for "Amount of Work Required" to .96 for "Liking for the Teacher" --- more than adequate for the kind of group measurements at issue.

The scale is applicable to grades 4-12. Note that norms are not required, since the items are scaled.

It is published by Education Employers Tests and Services Associates Publishing Company, 120 Detzel Place, Cincinnati, 19, Ohio.

Sister Amatora has published three articles on her further research on the scale (1950, 1952, 1954).

Ratings at the Secondary Level

In an experiment to test the hypothesis that prospective teachers could be taught to change their classroom behavior (Ward, Remmers, and Schmalzried, 1941) when informed of their pupils' attitudes as measured by the Purdue Rating Scale for Instructors, 40 practice teachers under Ward's supervision were rated by their students one month after they began teaching and again at the end of the semester. After the first rating Professor Ward conferred with each student-teacher concerning the teacher's general standing and specific strengths and weaknesses revealed by the first rating. Each student-teacher, but not his students, knew that the ratings would be repeated at the end of the semester.

The effect of the first ratings and conferences was shown by the differences between the two ratings. Only one

of the forty student teachers failed to gain in average rating. The average gain in all traits for the entire 40 was highly significant. The greater gains were for ratings for "self-reliance and confidence" and "sense of proportion and humor." The diagnostic and remedial value of the ratings was reflected in the relatively greater gains in the two traits in which student-teachers are probably most deficient, and in the general gains.

Four Decades of Student Ratings of Teachers at Purdue University and Other Institutions

About 40 years ago I said to one of my colleagues - the late George Brandenburg - "We're always rating students. I wonder what we'd get if we reversed the procedure and let them rate us?"

That gleam in my eye led to the conception of the Purdue Rating Scale for Instructors, and made machine scorable and later extended with P. C. Baker beyond the ten characteristics dredged from the then copious literature on teacher traits related to teaching effectiveness. This extension included 16 characteristics of the classroom situation, adequacy of text, laboratory, tests, and the like. Last year an IBM card was designed to make the rating scale reusable.

After having experimented with the scale to satisfy myself of its reliability, I went to the then President Elliott, showed it to him, and proposed to him that he send a letter that I would prepare to the members of the teaching faculty inviting them to have themselves rated, and that he increase my budget to meet the cost of the program. The letter stipulated that no administrator should have any of the ratings. The teachers concerned could score them themselves or my division would do it for them. The scales would be placed in the University Library where each teacher would ask for the number he required. A tear slip of the bottom of the letter requested the number each would require. This I needed to estimate the quantity to be printed.

In making the printing estimate I allowed a safety factor of 100%. A little more than half of the faculty initially decided to have themselves rated. When the program got under way, however, what I as a putative psychologist should have, but didn't anticipate, happened. The students began to put on the pressure.

"What's the matter with Professor X? Is he afraid to be rated?"

When it was all over 94% of the faculty had "voluntarily" had themselves rated! It made pretty representative norms.

Resistance to Student Ratings

As Robert Burns' "To a Louse" has it,

"O wad some power the giftie gie us
To see ourselves as ithers see us.
It wad free mony a blunder free us
And foolish motion."

But to face ourselves as others see us is threatening and takes some courage. It is likely to create a conflict with our self-perception and entail not a little cognitive dissonance.

Many criticisms of the whole business came to me, both in conversation and by letter. If the criticism contained a testable hypothesis, we set up the required research to test it. Following are some of the major conclusions drawn from these researches.

Conclusions

Rennumers (14) arrives at the following conclusions after a survey of investigations and studies related to student rating of instruction.

1. A considerable number of those who have used student ratings believe this procedure is useful for facilitating the educational process.
2. Knowledge of student opinions and attitudes leads to the improvement of the teacher's personality and educational procedures.
3. There is some evidence that student opinion is positively related to achievement as measured by examination of students.
4. If twenty-five or more student ratings are averaged they have as much reliability as do the better educational and mental tests at present available.
5. Grades of students are not, in general, related to their ratings of the teacher.

6. While the effect on student ratings of a generalized attitude (halo effect) toward the teacher has not been isolated, it apparently does not exist to an extent sufficient to invalidate the ratings of separate aspects of teaching methods and of the course. Evidence indicates that students discriminate reliably for different aspects of the teacher's personality and of the course, and between different instructors and courses.

7. There is evidence showing that little if any relationship exists between student ratings of teachers and the judged difficulty of the course.

8. In a given institution there exist wide and important departmental differences in effectiveness of teaching as judged by student opinion.

9. The sex of the student raters bears little or no relationship to their ratings of teachers.

10. The cost in time and money of obtaining student opinion is low. In fact, it is considerably lower than the administration of a typical standardized educational test of some comprehensiveness.

11. Popularity in extra-class activities of the teacher is probably not appreciably related to student ratings of that teacher. For instance, in a certain department of Purdue University the ratings of all instructors were uniformly high. However, teachers in this department are not at all prominent for their extra-class activities.

12. No research has been published invalidating the use of student opinion as one criterion of teaching effectiveness.

13. A positive relationship ($r=.24$) exists between student achievement and ratings awarded after initial ability has been partialled out.

14. Teachers with less than five years experience tend to be rated lower than teachers with more than 8 years experience.

15. The sex of the instructor has no effect on the ratings received.

16. The year in school of the rater has no effect on the ratings given, except that ratings by graduate students tend to be a bit higher than those by undergraduates.

17. Alumni of 10 years standing tend to rate their former instructors in the same way as do the present students of those instructors.

18. Students are more favorable to student ratings than instructors but more instructors have noticed improvement in their teaching as a result of student ratings than the students.

Graphic and numerical rating scales usually when used result in negatively skewed distributions, a function of a tendency of raters to be lenient in their recorded perceptions. To counteract this tendency we developed at Purdue University a "forced choice" instrument. Its application and validation was the doctoral research of John H. Snedeker. He obtained from 471 seniors 7,065 descriptive behavioral items which were finally distilled into two equivalent forms.

Administration time is about 10-15 minutes. For a representative sample of 112 teachers the reliability (internal consistency, Horst formula) was .96. The stability coefficient (Form A vs. B) three weeks interval was a Pearson r of .95.

The limitations of the forced choice rating scale are that (1) it yields no diagnostic information but only a single score and (2) the use of rating results in counseling rates will seriously risk invalidating the scale by making known the discriminating items.

Let me return once more to possible administrative use of student ratings. In a recent issue of the Saturday Review the distinguished former President of Brown University in an article under the title "Publish or Perish" has this to say:

"Under the best of circumstances, estimating the character and quality of a man's teaching is enormously difficult. In the short run, student opinion regarding teaching ability is virtually worthless. In the long run, student judgement about a teacher is nearly always sound, but early decision regarding tenure must be made before the long-run view has had an opportunity to manifest itself."

This statement in the light of our research results, is an ex cathedra assertion without factual basis and refuted by careful research. A study designed to compare some 114 teachers' ratings by alumni of ten years standing with ratings by on-campus undergraduates showed remarkable agreement.

References

- Bowman, E. C. Pupil ratings of student teachers. Education-
al Administration and Supervision, 1934, 20, 141-147.
- Kelley, T. L. The Influence of Nurture upon Individual Dif-
ferences, New York: The Macmillan Co., 1926.
- Remmers, H. H. and Gage, N. L. Educational Measurement and
Evaluation, Revised Ed., New York: Harper & Bros.,
1955, pp. 493-497.
- Tschechtelin, Sister M. Amatora, Hipskind, Sister M. John
Frances, and Remmers, H. H. Measuring the attitudes
of elementary school children toward their teachers,
Journal of Educational Psychology, 1940, 31, 195-203.
- Ward, W. D., Remmers, H. H. and Schmalzried, N. T. The
training of teacher-personality by means of student-
ratings, School and Society, 1941, 53, 189-93.
- Wriston, Henry M. Publish or perish, Saturday Review July
17, 1965, p. 59.