

Library School Instructor Evaluation

Since 1960 students have been asked to evaluate all courses taken in the Drexel library school. Results of the evaluations are tabulated and implications are discussed; they are also compared with similar studies in other disciplines and at other institutions. Summer school courses were better liked than those taken during the year; women instructors scored higher than men; courses in specialized library work ranked higher than others.

IN 1960 THE DREXEL INSTITUTE OF TECHNOLOGY graduate school of library science initiated a program of evaluation of faculty members by students. Student ratings are obtained for each section taught by full-time instructors during their first three quarters at Drexel, and for the first four courses taught by part-time instructors. Normally the evaluations are obtained in the last class meeting before the final examination.

The evaluations are intended to serve three purposes. Primarily, it is hoped that the ratings will aid the instructor in improving his teaching methods by pointing out specific areas needing improvement and by revealing to the instructor his students' reactions to him, both positive and negative, since "only by accident will the teaching of a man ignorant of the reaction of his class be effective."¹ Second, the ratings assist the administration in judging faculty members' effectiveness. A third purpose is to give students a voice in school administration.

¹W. R. Wilson, "Students Rating Teachers," *Journal of Higher Education*, III (February 1932), 79.

Miss Donaldson is Research Assistant and Dr. Harvey is Dean of the Graduate School of Library Science, Drexel Institute of Technology.

Since the 1920's, and especially in the last two decades, student evaluation of instructors has received increasing attention from educators and researchers. In general, this research has shown student evaluation to be reliable and valid. In his survey of the subject in the *Handbook of Research on Teaching*, H. H. Remmers cites various studies which have shown that student ratings were not appreciably influenced by the difficulty of the course, by the halo effect, by the grades given to raters, or by the instructor's popularity in extracurricular activities. Remmers cites further studies showing when twenty-five more ratings were averaged they were as reliable as the better mental and education tests, and when alumni graduated ten years earlier were asked to rate their college instructors, their ratings agreed substantially with those of the same instructors by students currently enrolled.²

The use of student evaluation is apparently widespread: in 1960 Stecklein cited evidence that 320 colleges and universities in the United States had used student ratings.³ Despite the amount of attention given student ratings on the undergraduate level, however, this subject has received little at-

²N. L. Gage, ed., *Handbook of Research on Teaching* (Chicago: Rand McNally, 1963), p. 367-68.

³*Ibid.*, p. 368.

tention at the graduate level and apparently none in library science. Yet it would seem that ratings might play a particularly important role in library education since, like instructors in other professional schools, most library school instructors have been trained primarily as professional practitioners rather than as teachers, and sometimes "the teachers drawn from the ranks of the profession are inadequately prepared for teaching."⁴ For the instructor with little experience or training in teaching methods, student ratings can point out areas on which he must concentrate to increase his effectiveness.

With the hope that a study of the results of student evaluations might provide insights into the problems of teaching library science on the graduate level, a study was made of the five-year Drexel evaluation program. The data upon which the study was based were limited in several aspects: in the first place, since ratings were generally obtained only for part-time and full-time instructors while they were still relatively new to the job, they do not necessarily provide a representative sample of all sections taught; second, the form of rating sheet has undergone slight modifications;

and third, in a few cases not all of the data was recorded on the master sheets.

The evaluation forms filled out anonymously by the students consisted of a list of attributes generally recognized to be associated with effective teaching. For each attribute students were asked to rate the instructor on a four-point scale—*excellent*, *good*, *fair*, and *poor*. The rating unit was the individual class section, so if an instructor taught three class sections in a given quarter he received three separate ratings. In compiling the scores for each section a master sheet was made which recorded the number of times each point on the grading scale was marked for each attribute. The total number of marks for each point on the grade scale was then figured. The final rating score was the percentage of all the marks which were at the *excellent* and *good* points of the scale. For instance, if thirty students in a section made a total of 200 evaluation decisions, 180 of which were at the *excellent* and *good* points of the scale, the instructor's rating for that section was 90 per cent.

Also recorded on the master sheet was a list of all voluntary student comments and the rank of that section compared with all other sections rated that quarter. At the end of the quarter—after all course grades had been turned in—a re-

⁴Thelma Eaton, "Who is a Good Library School Teacher?" *Improving College and University Teaching*, II (May 1954), 26.

TABLE 1. FREQUENCY DISTRIBUTION OF STUDENT RATINGS

SCORES	ALL RATINGS		RATINGS FOR PART-TIME INSTRUCTORS		RATINGS FOR FULL-TIME INSTRUCTORS		RATINGS FOR WOMEN		RATINGS FOR MEN	
	No.	Per Cent	No.	Per Cent	No.	Per Cent	No.	Per Cent	No.	Per Cent
96-100	30	12.34	17	12.78	13	11.81	14	12.06	16	12.59
91-95	40	16.46	22	16.54	18	16.36	23	19.82	17	13.38
86-90	39	16.04	21	15.78	18	16.36	16	13.79	23	18.11
81-85	32	13.16	14	10.52	18	16.36	17	14.65	15	11.81
76-80	31	12.75	18	13.53	13	11.81	17	14.65	14	11.02
71-75	21	8.64	8	6.01	13	11.81	11	9.48	10	7.87
66-70	19	7.81	13	9.77	6	5.45	7	6.03	12	9.45
46-65	24	9.87	13	9.77	11	10.00	11	9.48	13	10.23
26-45	7	2.87	7	5.26	7	5.51
	243	100.00	133	100.00	110	100.00	116	100.00	127	100.00

port was given to the instructor on an evaluation form, and a copy was kept by the administration.

The study considered separate course ratings representing 243 sections taught by eighty-four instructors. In analyzing these ratings the following questions were considered.

1. *What was the range for individual instructors' scores?* The scores ranged from 28 per cent to 100 per cent with the mean being 81 per cent shown in Table 1. Twenty-nine per cent of the ratings were above the 90 per cent level and an equal number below 75 per cent. The widest range for the scores of an individual instructor was 47 points, from 42 per cent to 89 per cent; however, the range of most instructors was considerably narrower, with the average range for instructors rated twice being seventeen points and the average for those rated three or more times being twenty-five points. Only one instructor was included in both the best twenty-five and the poorest twenty-five ratings.

2. *Did a pattern of movement exist in the scores of individual instructors or the ratings as a whole?* At Colorado State College of Education a survey of ratings taken over a period of years found that most instructors made significant improvements with successive evaluations.⁵ Riley cites three other studies which found a definite improvement in the performance of faculty members.⁶ At Drexel, however, no such trend was apparent, and while some of the instructors did receive successively higher ratings, as many received successively lower ratings or had such varying scores as to show no pattern of movement. Therefore, while Drexel faculty members on a whole learned their teaching weaknesses their resulting teaching

seemed no more—or less—effective than before. Of course, this generalization hides the group which did improve and the group which was already very high. As a matter of fact, the mean score of 81 per cent *excellent* and good was itself a quite satisfactory score. But minorities scored low or else had declining scores and were not invited to teach again.

TABLE 2. MEAN SCORES OF FACULTY RATINGS BY YEARS AND QUARTERS

Year:	N	MEAN SCORE
1960-61	26	81.2
1961-62	46	80.5
1962-63	43	83.5
1963-64	60	80.5
1964-65	56	76.2
<i>Quarter:</i>		
Fall	52	79.0
Winter	54	78.8
Spring	43	78.3
Summer	77	82.5

Table 2 shows that the ratings as a whole have recently shown a downward trend. While the average ratings for the first three years ranged from 80.5 per cent to 83.5 per cent, with the academic years 1963-64 and 1964-65 the average score declined to 76.2 per cent. Whether this trend represented an increasingly critical and able student body or a somewhat less effective faculty is not clear.

3. *Was there any connection between favorability of ratings and time of year?* In the Colorado State College of Education survey summer school courses were rated higher than those taught during the regular college year.⁷ The same trend was evident at Drexel where summer quarter course ratings averaged 82.5 per cent while fall, winter, and spring quarter averages were 79 per cent, 78.8 per cent, and 78.3 per cent respectively.

⁵ W. D. Armentrout, "Relation of Class Size and Other Factors to Student Ratings of Teachers," *Journal of Teacher Education*, I (June 1950), 102.

⁶ John W. Riley, Pryce F. Ryan, and Marcia Lifshitz, *The Student Looks at His Teacher* (New Brunswick: Rutgers University Press, 1950), p. 31.

⁷ "Relation of Class Size and Other Factors . . .," *op. cit.*, p. 102.

TABLE 3. COMPARISON OF PART-TIME AND FULL-TIME, AND OF MALE AND FEMALE INSTRUCTORS

	Mean Score	Per Cent of All Ratings	Per Cent of Highest 25 Ratings	Per Cent of Lowest 25 Ratings
<i>Ratings:</i>				
Full-Time Instructors	82.8	45	40	32
Part-Time Instructors	78.9	55	60	68
		100	100	100
<i>Ratings:</i>				
Male Instructors	79.2	52	52	64
Female Instructors	81.7	48	48	36
		100	100	100

More striking than a comparison of average scores, however, was the fact that while summer quarter ratings accounted for only 28 per cent of all the ratings they accounted for 68 per cent of the twenty-five highest ratings.

4. *Which groups of instructors scored highest?* A comparison was made of the scores received by women (48 per cent of all ratings) and by men (52 per cent). Table 3 shows the mean rating for women (81.7 per cent) to have been slightly but not significantly higher than that for men (79.2 per cent). Male instructors made up half of the top twenty-

five instructors but two-thirds of the bottom twenty-five instructors.

A similar comparison was made between part-time and full-time instructors. Again the difference was slight, with the mean rating for part-time being 78.9 per cent and for full-time being 82.2 per cent. Part-time faculty members tended somewhat more than full-time to scatter to either extreme. They represented 55 per cent of all ratings but had 60 per cent of the top twenty-five ratings and 68 per cent of the bottom twenty-five. In general, sex and part-time or full-time status appeared to have

TABLE 4. PERCENTAGE OF RATINGS FOR WHICH EACH ATTRIBUTE SCORED HIGHEST

ATTRIBUTE	Per Cent				
	All Ratings	Ratings for Part-Time Instructors	Ratings for Full-Time Instructors	Ratings for Male Instructors	Ratings for Female Instructors
Effectiveness in putting subject across	3	3	4	2	5
Usual preparation for class	16	9	23	14	18
Use of examples	4	4	4	3	5
Stimulation of thought	3	4	3	3	4
Tolerance toward student difference of opinion	25	33	16	30	19
Spirit of helpfulness in and out of class	25	24	27	32	18
Has motivated me to do my best work	1	2	0	0	2
Balance of lectures, class discussion, student reports	0	0	0	0	0
Stimulates exchange of ideas	1	1	0	0	1
Assignments reasonable in length	2	3	1	3	1
Knowledge of and familiarity with subject	50	59	39	48	52
Organization of material	5	3	7	3	7

TABLE 5. PERCENTAGE OF RATINGS FOR WHICH EACH ATTRIBUTE SCORED LOWEST

ATTRIBUTE	PER CENT				
	All Ratings	Ratings of Part-Time Instructors	Ratings of Full-Time Instructors	Ratings of Male Instructors	Ratings of Female Instructors
Effectiveness in putting subject across	18	25	10	21	14
Usual preparation for class	9	11	6	12	4
Use of examples	0	0	0	0	0
Stimulation of thought	29	29	28	35	21
Tolerance toward student difference of opinion	11	5	18	4	20
Spirit of helpfulness in and out of class	3	4	2	2	4
Has motivated me to do my best work	5	6	3	5	4
Balance of lectures, class discussion, student reports	6	8	3	7	4
Stimulates exchange of ideas	4	2	6	3	5
Assignments reasonable in length	7	8	5	3	10
Knowledge of and familiarity with subject	1	1	1	0	2
Organization of material	31	33	28	34	26

little bearing on teaching effectiveness, but the interesting clusters at extremes are hard to explain.

5. *In what characteristics were instructors rated highest? Lowest?* The strongest characteristic was "knowledge of and familiarity with subject," which was a strong point on half of the ratings. "Tolerance toward student difference of opinion" and "spirit of helpfulness in and out of class" were each strong points on 25 per cent of the courses. Table 4 shows no other characteristics to be rated highest on more than 10 per cent of the courses.

The weakest characteristics were "organization of material" (31 per cent), "stimulation of thought" (29 per cent), and "effectiveness in putting subject across" (18 per cent) as seen in Table 5.

It is interesting to compare Drexel strong and weak points with the results of a large-scale rating program carried out among undergraduates at Brooklyn College and reported by Riley, Ryan, and Lifshitz in *The Student Looks at His Teacher*. In the Brooklyn study, as at Drexel, the highest scoring characteristic

concerned knowledge of subject matter. Similarly, at Brooklyn the poorest scoring characteristic was "encouragement of thinking," and at Drexel "stimulation of thought" was one of the weakest points. Despite these correspondences, however, there was a significant difference between Drexel and Brooklyn instructors on "organization of subject material"—the third best characteristic of ten at Brooklyn, but the weakest at Drexel.⁸

There were several differences in the weakest and strongest characteristics by sex and by full-time status of faculty members. For instance, while almost one-fourth of the full-time instructor ratings were strong in "usual preparation for class," this was a strong point for only one-eleventh of the part-time instructors who usually had full-time jobs elsewhere in addition to their Drexel teaching. On the other hand, ratings for part-time instructors were 50 per cent higher than for full-time on "knowledge of and familiarity with subject." Students also considered part-time instruc-

⁸ Riley, Ryan, and Lifshitz, *op. cit.*, p. 82.

TABLE 6. RATINGS OF TYPES OF COURSES

Type of Course	Mean Score	Per Cent of All Ratings	Per Cent of Top 25 Ratings	Per Cent of Low 25 Ratings
Reference	81.0	19	20	24
Library Materials	78.9	15	4	12
Cataloging	79.8	12	8	12
Administration	79.1	11	12	8
Special Types of Service	82.3	14	28	16
Backgrounds	77.0	12	4	20
Children's Work	88.6	10	20	0
Information Science	79.8	7	4	8
		100	100	100
Basic Required	79.2	29	12	28
Intermediate Required	79.5	25	12	28
Electives	82.0	38	72	36
Information Science	79.8	7	4	8
		100	100	100

tors more tolerant of student difference of opinion, on a two-to-one ratio. Still another significant difference between part-time and full-time instructors lay in the area of "effectiveness in putting subject across" which was a strong point for one-fourth of the part-time instructors but for only one-tenth of the full-time instructors.

An interesting difference between the ratings of male and female instructors was in "tolerance toward student difference of opinion." This attribute occurred as a weak point on ratings of female instructors significantly more often than on ratings of males.

6. *Were electives rated higher than required courses?* Although it may seem natural for students to have been more favorably inclined toward the courses they elected to take than toward required courses, the studies done at Brooklyn College⁹ and at Colorado State College of Education¹⁰ found no important difference here. At Drexel, as shown in Table 6, elective courses had a slight but not significantly higher mean than basic and intermediate required courses.

⁹ *Ibid.*, p. 87.

¹⁰ "Relation of Class Size and Other Factors . . .," *op. cit.*, p. 102.

Required courses, however, accounted for 54 per cent of all the ratings but only 24 per cent of the highest twenty-five ratings.

7. *Which courses were most highly rated?* When considering the scores of groups of courses the most obvious pattern was the high ratings given to courses dealing with library service to children and young people in school and public libraries. The seven courses in this category were offered twenty-five times for an average score of 88.5 per cent, eight points higher than the overall average. Also, special types of library service, such as medical, law, special, college, etc., were unusually well represented among the top twenty-five courses. It is possible that the somewhat more favorable ratings given school and children's librarianship courses were related to the similarly favorable ratings given in summer quarters, since during summer quarters the percentage of students and of courses in school and children's librarianship is relatively high.

For single courses, the highest score for a required course rated more than ten times was for "Basic Reference Materials" with a mean score of 86 per cent. The lowest mean score for such a

TABLE 7. 25 HIGHEST RATINGS

Instructor	Part-Time or Full-Time	Male or Female	Course	Elective or Required	Class Size	Year	Quarter	Rating-Per Cent
A	F	F	Advanced Selective Cataloging	E	6			100
C	P	M	Rare Book Librarianship	E	9	63	Summer	100
D	P	F	Workshop	E	17	62	Summer	100
E	F	F	Adult Education and the Library	E	14	61	Spring	100
G	P	F	Advanced Reference	E	4	60	Summer	100
S	P	M	College and University Library Service	E	9			100
T	F	M	Methods of Research in Librarianship	E	4	60	Summer	100
B	P	F	Selection of Library Materials	R	19			99
K	P	F	Folk Literature and Its Oral Presentation	E	11	63	Summer	99
Q	P	F	Library Methods Analysis	E	7	62	Summer	99
DD	P	F	Selection of Library Methods for Children	E	20	61	Summer	99
H	P	M	Government Publications	E	22	64	Fall	98
I	F	M	Special Library Service	E	13	61	Summer	98
I	F	M	Special Library Service	E	10	63	Winter	98
CC	P	M	Integration of Science Information Systems	E	5	64	Summer	98
M	P	M	Introduction to Library Services	R	22	61	Summer	98
W	P	M	College and University Library Service	E	23	62	Spring	98
BB	P	F	Selection of Library Materials for Children	E	14	63	Summer	98
A	F	F	Cataloging and Classification	R	10			97
G	P	F	Advanced Reference	E	8	64	Fall	97
J	F	F	Selection of Library Materials for Young People	E	21			97
I	F	M	Introduction to Library Services	R	22	61	Summer	97
I	F	M	Reference in Science and Technology	R	24			97
I	F	M	Reference in Science and Technology	R	26	63	Winter	97
N	P	F	Adult Education and the Library	E	15	62	Summer	97

course was the 74 per cent received by "Library in Society," a required course on the history and sociology of libraries.

8. *Is there any relationship between class size and ratings?* Although class size was not a significant factor in the ratings at Brooklyn¹¹ and Colorado,¹² small classes at Drexel were rated more favorably than large ones, as Tables 7 and 8 show. The average class size at Drexel was twenty, but the average class size for the highest twenty-five ratings was only 14.2. It should be noted, however, that the average class size for the

twenty-five lowest courses—18.76—was also somewhat lower than the over-all average. The mean score for classes having ten or fewer students was 86 per cent, compared to the over-all average of 81 per cent. While small classes comprised only 10 per cent of all ratings, they made up 40 per cent of the twenty-five highest ratings. Furthermore, of all small classes, more than half received scores of 90 per cent or above.

Library science instructors are faced with many teaching problems. In addition to the fact—mentioned earlier—that many of them have had little or no training in teaching methods, they must teach classes made up of students with

¹¹ Riley, Ryan, and Lifshitz, *op. cit.*, p. 87.

¹² "Relation of Class Size and Other Factors . . .," *op. cit.*, p. 102.

TABLE 8. 25 LOWEST RATED COURSES

Instructor	Part-Time or Full-Time	Male or Female	Course	Elective or Required	Class Size	Year	Quarter	Rating—Per Cent
V	P	M	Introduction to Cataloging and Classification	R	8	61	Fall	28
F	P	M	Audio-Visual Materials	E	18	62	Spring	34
P	P	M	College and University Library Service	E	23	61	Summer	36
U	P	M	College and University Library Service	E	23	37
P	P	M	Selection of Library Materials	R	28	61	Summer	37
O	P	M	Library Methods Analysis	E	13	65	Winter	39
JJ	P	M	Selection of Library Materials	R	15	62	Summer	42
FF	P	F	Reference in Science and Technology	R	12	64	Spring	46
FF	P	F	Reference in Science and Technology	R	24	64	Winter	48
GG	P	M	Reference in Social Sciences	R	25	64	Fall	48
J	P	M	Science Lit. Searching and Abstracting	E	21	49
HH	F	F	History of Books and Printing	R	9	65	Winter	51
R	P	F	Introduction of Cataloging and Classification	R	33	62	Summer	52
HH	F	F	Reference in Social Sciences	R	18	64	Fall	53
II	P	M	College and University Library Service	E	31	65	Spring	54
HH	F	F	Library Administration	R	35	65	Winter	56
AA	P	F	Search Strategy	E	12	62	Fall	56
X	P	M	Methods of Research in Librarianship	E	19	63	Winter	56
Z	F	M	Reference in Humanities	R	26	64	Spring	57
EE	P	M	Law Librarianship	E	17	62	Spring	58
T	F	M	Library in Society	R	7	58
T	F	M	Library in Society	R	13	64	Winter	59
L	F	F	Introduction to Cataloging and Classification	R	17	65	Spring	59
HH	F	F	Methods of Research in Librarianship	E	11	65	Winter	60
Y	P	M	Advanced Reference	E	11	64	Summer	62

widely varying backgrounds, library experiences, and goals in librarianship. Furthermore, there is the constant problem of achieving a happy balance between theory and practice. Add to this the lack of adequate textbooks and it seems clear that the task facing the li-

brary science instructor is not an easy one. At Drexel it is felt that the student rating program is an important method of helping the instructor do an effective job. While ways of improving the rating sheets are constantly being sought the program itself has proven successful. ■■

