Higher Education 2 (1973) 439–446

THE DEPARTMENTAL RATING GAME: MEASURE OF QUANTITY OR QUALITY?

CHARLES F. ELTON and SAM A. RODGERS

University of Kentucky
Lexington, Kentucky

ABSTRACT

What constitutes quality in graduate education? The most widely accepted definition has been proposed by Cartter who asked recognized scholars to rate departmental excellence in 30 disciplines. Cartter's instructions to the raters could have allowed the influence of a "halo effect" to operate. This is an error in rating which is produced when the particular characteristics being rated are contaminated by the rater's notion of the general worthiness of the object being rated. This study demonstrated that the halo effect related to size variables occurred in the Cartter study.

Data were collected from public sources for each department of mathematics, physics, chemistry, and geology rated in the Cartter study as "extremely attractive," "attractive," "acceptable plus" and for a random sample of "less than acceptable plus" departments. These data consisted of the following size variables: (1) number of areas of specialization within a department; (2) number of faculty; (3) number of Ph.D. degrees awarded between 1960–64; (4) number of full-time students; (5) number of first year students; and (6) ratio of part-time to full-time students.

Tests of statistical significance indicated that these variables differentiated the departmental ratings beyond chance expectation. A graphic illustration is provided for each discipline showing the relationships between the size measures and the mean departmental ratings.

Implications of these findings are that measures of size ought not to be confused with measures of quality and that the development of measures of quality is a matter of urgent priority.

Quality in graduate education, like love, appears to exist in the eye of the beholder. Experienced writers in the area of higher education do not always agree on how to recognize quality (Dressel *et al.*, 1970; Woodring, 1968). Dressel and his colleagues avoid discussion of the essence of quality and emphasize the difficulties in separating measures of quality from quantity (pp. 134–36). Woodring, however, states, "It is somewhat easier to assess the quality of graduate schools than of undergraduate liberal arts colleges because the goals of graduate education are more sharply defined and the number of institutions is much smaller" (pp. 50–51). This author then describes the Cartter (1966) study as an example of an investigation of quality in graduate education.

Cartter sought ratings by established scholars of faculty and program excellence in various fields of study at the graduate level. His instructions to his faculty raters were as follows:

> How would you rate the institutions below if you were selecting a graduate school to work for a doctorate in your field today? Take into account the accessibility of faculty and their scholarly competence, curricula, educational and research facilities, the quality of graduate students and other factors which contribute to the *significance of the doctoral program.*

Most recently, Roose and Andersen (1971), in an expanded follow-up of the Cartter study, gave the same instructions to their panels of faculty raters.

It is probable that these instructions allowed the results to be contaminated by a halo effect. This is an error in rating which is produced when the particular characteristics being rated are contaminated by the rater's notion of the general worthiness of the object being rated. That is, there are six different items which are explicitly referred to in the rating instructions, i.e., accessibility of faculty, scholarly competence of faculty, curricula, educational facilities, research facilities, and the quality of graduate students. The addition of "other factors" allowed each rater to provide his own definition of items of importance. It has been established for some time that the validity of a rating is determined by several characteristics, one of which is the degree to which it is specific rather than general (Guilford, 1954). The more general the characteristic being rated, the greater is the possibility of error, since raters will base their judgments on different aspects of the composite.

This study will demonstrate that the institutional ratings of physical science departments appearing in the Cartter study can be predicted by the use of simple data in the public domain and which incorporate variable related to size.

Graham (1965) has provided for aspiring graduate students some objective data concerning graduate programs. From this volume, the following variables were arbitrarily selected: (1) number of areas of specialization within a department; (2) number of faculty; (3) number of Ph.D. degrees awarded between 1960–64; (4) number of full-time students; (5) number of first year students; and (6) ratio of part-time to full-time students. These data were collected for the following physical science departments rated in the Cartter study: mathematics, physics, chemistry and geology. The departments of astronomy within the physical science area were not included because of the small number of departments which were rated. Although the publication date of the Graham book is 1965 and that of the Cartter study is 1966, it was assumed that the data contained in the former fairly represents the departments during the period of time covered by the Cartter investigation.

The six size variables described above are used as predictor variables in four separate stepwise multiple discriminant analyses in which the departments rated in the Cartter study as "extremely attractive," "attractive," and "acceptable plus" were the dependent variables. An additional dependent variable for each discipline consisted of a random selection of departments rated "less than acceptable plus" which were taken from Appendix E in the Cartter study. Hereinafter, these categories will be referred to as "A" (extremely attractive), "B" (attractive), "C" (acceptable plus) and "D" (less than acceptable plus) departments.

The discriminant analyses were used to demonstrate that the arbitrary selection of predictor variables did indeed differentiate the

TABLE I

Predicted Classification of Departments

Mathematics

Category	"A"	"B"	"C"	"D"	N	Percent Correct
"A"	6	1	1	0	8	75
"B"	1	5	3	2	11	45
"C"	2	2	14	5	23	61
"D"	0	0	5	8	13	62

Overall % = 60; N = 55

Physics

Category	"A"	"B"	"C"	"D"	N	Percent Correct
"A"	6	2	0	0	8	75
"B"	1	9	1	0	11	82
"C"	0	5	18	5	28	64
"D"	0	0	1	13	14	93

Overall % = 75; N = 61

Chemistry

Category	"A"	"B"	"C"	"D"	N	Percent Correct
"A"	6	2	1	0	9	67
"B"	3	9	0	1	13	69
"C"	2	6	21	8	37	57
"D"	0	0	3	9	12	75

Overall % = 63; N = 71

Geology

Category	"A"	"B"	"C"	"D"	N	Percent Correct
"A"	5	1	0	1	7	71
"B"	1	3	4	1	9	33
"C"	1	2	11	4	18	61
"D"	0	0	1	12	13	92

Overall % = 66; N = 47

departments in the same way that the Cartter ratings did. Subsequently, the relationships among the predictor variables were shown by the spatial configuration technique developed and described by Cole and Cole (1970).

Results

In each discipline, the discriminant analysis provided a statistically significant differentiation ($p < 0.01$) among the "A", "B", "C", and "D" rated departments. The effectiveness of the predictors in classifying the physical science departments according to Cartter's categories is given in Table I, in which the "hit" rate is indicated by the numbers in bold face type.

The overall efficiency of prediction among physical science depart-

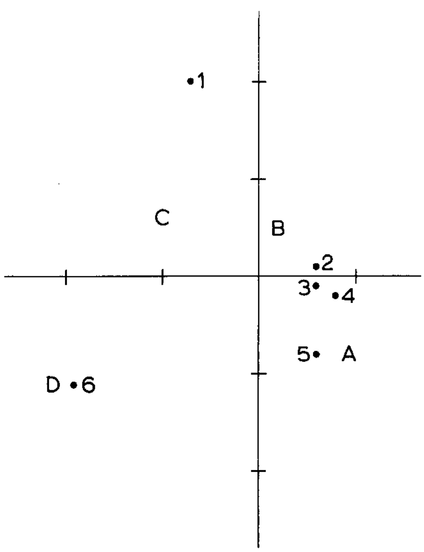

Fig. 1. Spatial configuration of six predictor variables (Mathematics).

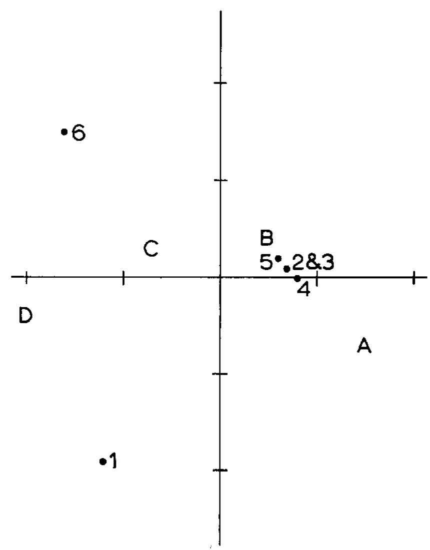

Fig. 2. Spatial configuration of six predictor variables (Physics).

1 = Number of areas of specialization;
2 = Number of faculty;
3 = Number of Ph.D.'s awarded 1960–64;
4 = Number of full-time students;
5 = Number of first year students;
6 = Ratio of part-time to full-time students;
A = Mean profile point for "A" departments;
B = Mean profile point for "B" departments;
C = Mean profile point for "C" departments;
D = Mean profile point for "D" departments.

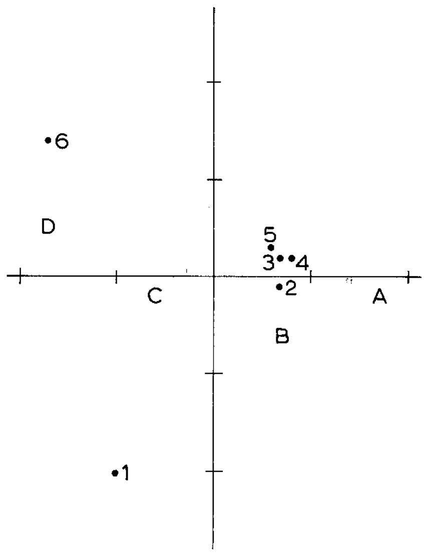

Fig. 3. Spatial configuration of six predictor variables (Chemistry).

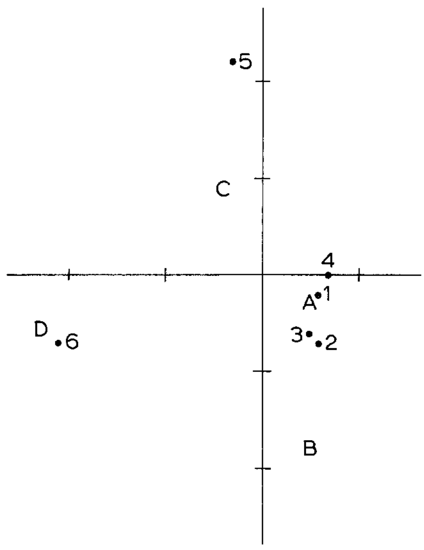

Fig. 4. Spatial configuration of six predictor variables (Geology).

ments ranged from a high of 75% in physics to a low of 60% in mathematics. Generally, the best predictions occurred in the "A" and "D" rated departments.

Figures 1–4 include the spatial relationships among the predictor variables and the mean predictor profiles for all "A", "B", "C", and "D" rated departments in mathematics, physics, chemistry, and geology, respectively.

The spatial configurations of mathematics and physics are strikingly similar (Figures 1 and 2). "A" and "B" mean profiles are most influenced by number of faculty (variable 2); number of Ph.D. 's awarded from 1960 through 1964 (variable 3); number of full-time students (variable 4); and number of first year students (variable 5). "C" and "D" mean profiles are most influenced by number of areas of specialization within a department (variable 1) and ratio of part-time to full-time students (variable 6).

"A" and "B" mean profiles in both chemistry and geology are located in the same quadrant (Figures 3 and 4). The mean profiles of chemistry "A" and "B" departments are, like those in mathematics and physics, more subject to the combined influences of variables 2, 3, 4, and 5 than are those in geology. The chemistry "C" and "D" mean profiles are also influenced by variables 1 and 6 whereas those in geology are dominated by variables 5 and 6.

Discussion

It seems clear that a halo effect related to size variables influenced the faculty ratings of physical science departments. Furthermore, the magnitude of this error is such that it is doubtful whether the faculty rating method offers any substantial advantage over the use of public data in the assessment of departmental reputation.

The mean departmental profiles of "A" and "B" departments are found to be consistenly related to the following three variables: number of faculty (variable 2) number of Ph.D.'s awarded between 1960–64 (variable 3); and number of full-time students (variable 4). These three variables are highly intercorrelated and this relationship is retained when the variables are projected onto a plane (Figures 1–4). In short, "extremely attractive" and "attractive" mean ratings are characterized by large size. This finding agrees with the results reported by Wispé (1969) which indicated that productivity in psychology departments, as measured by publication criteria, was related to size during the 1960 decade.

Mean departmental profiles of "C" and "D" departments, with the exception of geology, are influenced by two variables: number of areas of specialization within the department (variable 1) and ratio of part-time to full-time students (variable 6). Low rated departments, compared to those with high ratings, may be characterized as offering a larger number of departmental options to a larger number of part-time students.

To the prospective graduate student seeking to make an informed choice of a physical science graduate program characterized by quality, the better part of valor would seem to consist of seeking admission to one of the larger departments. To faculty concerned about the quality of graduate study in physical science, the better part of wisdom consists of recognizing the fact that although large size is not necessarily incongruent with quality, neither is it an absolute guarantee of quality.

Research studies which have a potential for influencing national policy are not common. The Cartter and Roose–Andersen studies possess this potential and, as a consequence, some governmental agencies have considered such data in policy statements (Scully, 1970a); some graduate deans have expressed fear as to how these results will be used (Scully, 1970b). The general implications for policy are highlighted by the following statement taken from *A Rating of Graduate Programs* (Roose and Andersen, 1971, p. 25).

> From the standpoint of national policy, consideration must be given to the possibility that in the future a more than sufficient supply of Ph.D.'s for most traditional uses can be trained in the

graduate programs of say, 50 or so top-rated institutions. Thus, to take only one illustration, Cartter has forecast that, given present trends, the supply of Ph.D.'s in 1980–81 will number 50,000. This supply compares with the forecast of demand for new doctorates for teaching in 1980–81 (assuming the maintenance of quality) of 7,400. According to 1966–67 data, the top 50 institutions in our 1969 ratings study awarded 66.8 percent of the nation's doctorates. Even if one assumes the percentage of doctorates awarded by these institutions falls to 40 percent by 1980–81, they would still produce over 20,000 Ph.D.'s, or almost three times the number needed for maintenance of quality in teaching.

It is not surprising therefore that analyses of the imbalance between the supply and demand of graduate students as well as reviews of national manpower requirements are prominent themes which dominate the literature on graduate education in America (Folger *et al.,* 1970; Scully, 1970c; Cartter, 1971). Furthermore, in a time of diminishing public support for higher education, it might be expected that some would propose allocating scarce federal and state funds to those top rated departments in which quality exists and to restrict graduate study in other departments (Higgins, 1971).

The issue raised in this study, however, is not the significance or necessity of national manpower studies; our findings illustrate the all-too-common tendency to measure size and equate it with quality. For example, the title of the Cartter 1966 study is *An Assessment of Quality in Graduate Education.* Our society is permeated with the simple notion that the biggest is the best, and it is unfortunate when our graduate programs are assessed entirely by the same yardstick.

The issue, of course, is: What constitutes quality? Perhaps a beginning consists of clearly recognizing that quality is a value judgment and that neither philosophers nor physical scientists are exempt from this starting point. Administrators and faculty might derive some suitable measures of quality from the description of an instructional model proposed by Glaser (1970). He proposes, among other conditions, that learning outcomes should be defined in terms of behavioral manifestations of competence. More importantly he suggests that a careful analysis be made of the characteristics of the learner before instruction begins. Once these characteristics are determined, alternative routes of learning experiences should be available so that the student can maximize his learning potential. Thus, given the present state of graduate study in the United States, an investigation of quality in graduate education might consist of an assessment of the extent to which departmental faculties are engaged in the following: determining the characteristics of the potential learner, devising alternative educational experiences which are related to the strengths and weaknesses of its students, and evaluating its students in terms of behavioral competencies.

In summary, this study found that departmental reputation in physical science, as it was assessed by the faculty raters in the Cartter study, is confounded with measures of size. Since the dividing line between "departmental reputation" and "quality"is, at present, mostly rhetorical, it appears desirable to investigate additional correlates of departmental effectiveness.

References

Cartter, A. M. (1966). *An Assessment of Quality in Graduate Education.* Washington, D. C.: American Council on Education.

Cartter, A. M. (1971). "Scientific Manpower for 1970–1985." *Science,* 172, pp. 132–140.

Cole, N. S. and Cole J. W. L. (1970). *An Analysis of Spatial Configuration and Its Application to Research in Higher Education,* Iowa City, Iowa: American College Testing Program.

Dressel, P. L., Johnson, C. and Marcus, P. M. (1970). *The Confidence Crisis.* San Francisco: Jossey-Bass.

Folger, J. K., Astin, Helen S. and Bayer, A. E. (1970) *Human Resources and Higher Education,* New York: Russell Sage Foundation.

Glaser, R. (1970) "Evaluation of Instruction and Changing Educational Models." In Wittrock, M. C. and Wiley, D. (eds.) *The Evaluation of Instruction,* New York: Holt, Rinehart and Winston.

Graham, Jane. (1965). *A Guide to Graduate Study.* 3rd edition. Washington, D. C.: American Council on Education.

Guilford, J. P. (1954). *Psychometric Methods,* New York: McGraw-Hill.

Higgins, A. S., (1971). "Quantity and Quality of Doctoral Overproduction". *Educational Record,* 52, 262–266.

Roose, K. D. and Andersen,C. S. (1970). *A Rating of Graduate Programs,* Washington, D. C.: American Council on Education.

Scully, M. G. (1970a). "U.S. Questions Policies Aiding Graduate Study". *The Chronicle of Higher Education,* December 14, p. 1–2.

Scully, M. G. (1970b). "Graduate Deans Criticize Forthcoming Assessment of Their Departments." *The Chronicle of Higher Education,* December 14, p. 3.

Scully, M. G. (1970c). "Graduate Study Growth Plans are Challenged". *The Chronicle of Higher Education,* October 26, p. 1.

Wispé, L. G. (1969). "The Bigger the Better: Productivity, Size and Turnover in a Sample of Psychology Departments". *American Psychologist,* 24, 662–668.

Woodring, P. (1968). *The Higher Learning in America: A Reassessment.* New York: McGraw-Hill.