Bulletin of The British Psychological Society (1985), **38**, 361–365 *Printed in Great Britain*

On the assessment of teaching effectiveness in British universities

J. Philippe Rushton and **Harry G. Murray**

Increasing interest is arising in Britain on how to teach psychology more effectively (e.g. Rose & Radford, 1984) and, indeed, more generally, on how to be a better teacher. The present report is spurred by Carrick-Smith's critical letter (*Bulletin*, May 1985), which implied there was 'no research which yielded any criteria for reliably identifying the "incompetent" teachers' (p. 157). While most people have a natural apprehension about the possibility of being negatively evaluated, or inappropriately judged, it should be noted that evaluation of academic performance can also serve the beneficial function of providing information for appropriate improvement. In this article we provide an overview of the method of evaluating teaching at our own university and review some of the literature connected therewith. Since feedback and performance improvement (learning) have always been mainstay topics in psychological research, it would seem most appropriate to study these processes directly in university classrooms as well as in the experimental laboratory!

Student rating-forms

The most common method of evaluating teaching in North American universities is formal student ratings, usually obtained by means of a standardized, objectively scored evaluation form on which students rate teacher characteristics such as clarity, rapport, preparation, and fairness in grading. Our own university, The University of Western Ontario, requires annual, end-of-course student evaluation of instructors in all courses. Ten items concerning various aspects of the teacher and course are rated on five-point scales, with the last item being a rating of the 'overall effectiveness' of the instructor. This evaluation form is shown in Table 1.

Recent surveys suggest that student ratings of instructors, on forms such as that shown in Table 1, are increasing rather than decreasing in frequency of use, and are easier to administer and less intrusive than other procedures such as videotaping of classroom teaching or classroom visitation by colleagues or department heads. At most North American universities (including our own), teaching is evaluated for either or both of two major purposes: (1) to provide faculty members with diagnostic feedback which, it is hoped, will lead to improved teaching performance and (2) to provide a basis for administrative decisions on faculty hiring, salary, contract renewal, tenure, and promotion.

The construct validity of student ratings

A general review of what is known about the reliability and validity of student instructional ratings was recently provided by Murray (1984). In regard to year-to-year reliability for the same instructor and course, there is good evidence for high stability of scores on scales such as that in Table 1. Reliabilities of 0·80 to 0·90 are typical, and do not vary enormously depending on the type of course being taught nor on whether alternate rating scales are used. Student ratings of classroom teaching also correlate 0·50 to 0·90 with comparable ratings made by supervisors, colleagues, alumni, and paid classroom observers, indicating that student perceptions of good and poor teaching are similar to those of more expert, more mature, and more neutral observers.

Of most importance in terms of validity is the evidence indicating at least a moderate positive correlation between student ratings of teaching and

Table 1. Teacher rating-form

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 material in a well-organized and coherent manner.
4. The instructor is able to explain difficult concepts in 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 the subject matter.
8. Considering inherent limitations of the course content the instructor is successful in 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?

objective measures of student achievement. As Murray (1984) discussed, most studies of this relationship have been done in the context of a multiple-section course with a common, objectively scored final examination, where it is possible to correlate mean student ratings with mean final examinations scores across instructors or class sections. Cohen (1981) reviewed 41 studies of this type and reported an average correlation of 0·43 between instructor ratings and student examination performance, indicating that highly rated teachers do in fact tend to foster higher levels of student learning than less highly rated colleagues.

Error variance does of course exist in student ratings, as in any other measure of human performance. For example, ratings tend to be higher in small classes than in large classes, and higher for instructors who assign high grades than for instructors who assign low grades. However, Marsh (1980) estimated that these extraneous factors account for no more than 12 to 14 per cent of the total variance in instructor mean ratings.

On the basis of many more findings than those reviewed above, most writers have concluded that student ratings of teaching are sufficiently reliable and valid to justify their use both as diagnostic feedback to the instructor and as one of several sources of information in administrative personnel decisions. As Murray (1984) has emphasized, however, it must be kept in mind that student ratings can provide information on only certain aspects of teaching, and even if proven totally reliable and valid for this limited purpose, must usually be supplemented by other sources of data. As a case in point, it is entirely possible that a faculty member who receives poor evaluations from students as a classroom teacher, might be outstanding in non-classroom teaching activities such as thesis supervision or development of instructional materials. Alternatively, it is conceivable that a faculty member who receives high ratings from students might be seriously lacking in knowledge of the subject matter, coverage or content, or academic standards. It is impossible to identify teachers such as these if we rely solely on student ratings.

Personality and teaching effectiveness

As with most performance measures, it appears that some people have an inherent advantage over others, and several studies have documented that the more extraverted, less anxious individual is seen as the more effective communicator in the classroom setting. Other traits predicting high teacher ratings include intelligence, leadership, liberality, objectivity, and supportiveness (Rushton *et al.*, 1983).

A recent study, however, demonstrated that the relationship between instructor personality and student evaluation of teaching is mediated by specific concrete classroom behaviours (Erdle *et al.*, 1985). First, some 21 personality traits on which faculty members had been rated by their peers were reduced to two factors accounting for 66 per cent of the variance. Factor 1: *achievement orientation*, was made up of traits such as dominance, ambition, leadership, intelligence, and endurance, while factor 2: *interpersonal orientation*, was made up of traits such as supportiveness, non-authoritarianism, non-defensiveness, and objectivity. Professors were assigned factor scores on the two underlying personality dimensions.

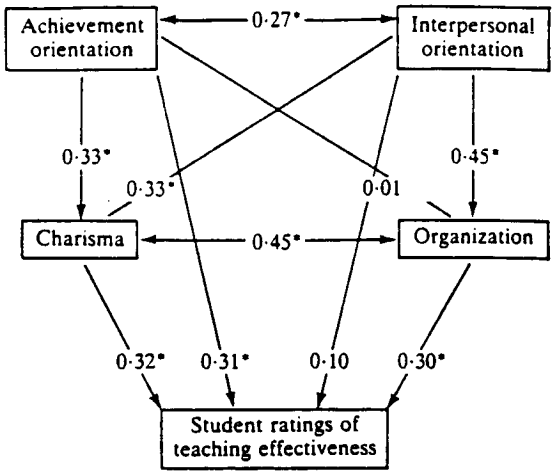
Second, some 95 specific classroom behaviours (gestures with hands, makes eye contact, defines new terms, gives preliminary overview of lecture, provides a logical sequence to lectures, shows interest in subject matter) were coded by an average of six trained classroom observers over an average of 20 hours of observation for each instructor. These classroom behaviours were also reduced to two factors, accounting for 51 per cent of the variance. Factor 1: *charisma*, was made up of such classroom behaviours as speaking expressively and showing interest in the subject matter. Factor 2: *organization*, consisted of items such as giving a preliminary overview of the lecture and giving organization to the lecture by the use of headings.

To test a causal model that classroom behaviour mediates the relationship between personality and teaching effectiveness, the path analysis shown in Fig. 1 was undertaken using the personality and classroom behaviour composites as predictors and student ratings of teaching effectiveness as the criterion. The path coefficients shown in the figure are standardized beta weights derived from a series of multiple regression analyses. Overall, the four predictors in the analysis accounted for 57 per cent of the variance in student ratings of teaching effectiveness, and 50 per cent of the relationship between personality and teaching effectiveness was mediated by classroom behaviour.

Can evaluation lead to improved teaching?

Given that good teaching is partly a matter of engaging in specific classroom behaviours, one might hope that it would be a relatively simple matter to improve teaching performance. Again Murray (1984) has reviewed the literature and has concluded that, indeed, evaluation feedback can and does lead to improvement in teaching effectiveness. Murray reviewed evidence from three different sources to support his conclusion (faculty opinion surveys, experimental manipulations of student feedback, and longitudinal field studies). The extent of such improvement seems to depend on, among other things, whether teaching evaluation is global or specific, whether evaluation is supplemented by expert consultation, and whether evaluation results are to be used in salary and promotion decisions.

It is also possible to directly train college teachers to communicate more effectively in the lecture hall.



*$p<.05$

Figure 1. Path analysis of composite personality dimensions and classroom behavior dimensions in relation to student ratings of teaching effectiveness. (Source: Erdle *et al.*, 1985).
*$P<0.05$.

Murray & Lawrence (1980) carried out such a study by training lecturers in acting techniques aimed at improving lecture skills. Twenty-four full-time faculty members from the Departments of Psychology, Sociology, and Physics at The University of Western Ontario served as participants. The 12 teachers in the experimental (training) group responded to a request for volunteer participants in a speech and drama programme aimed at improvement of lecture skills. The 12 teachers in the control (untrained) group were selected on the basis of a careful matching procedure. Teachers in the experimental group participated in a series of 20 two-hour group sessions on acting technique (a total of 40 hours of instruction) given by an experienced speech and drama instructor. The general objective of the programme was to train participants in acting skills and to suggest ways of applying these skills to classroom teaching. Specific activities in weekly sessions included breathing and voice exercises, reading of monologues and political speeches, acting out of short scenes from plays, practice in the use of vocal variation, pausing, eye contact, facial expression, and body movement to reinforce meaning, and videotaping of five-minute lecture segments with corrective feedback from the instructor and other participants during playback.

Classroom teaching performance of participating experimental and control teachers was measured by a specialized 42-item student rating-form administered by mail survey near the beginning (pre-test) and again near the end (post-test) of the 20-week series of acting lessons. To ensure procedural independence of measures, the students who provided pre-test ratings were different from those who provided post-test ratings. To check on the possibility that greater improvement in experimental teachers reflected a generalized placebo effect, the student rating-form included both 'target' items, which were expected to change as a result of training, and 'non-target' or control items, which were not expected to change. Also, students' ratings were traced back in time to determine whether experimental or control teachers had shown improvement prior to training. The results demonstrated that speech and drama training produced significant improvement in target teaching behaviours, but not in non-target behaviours. This effect was most marked for specific behavioural items such as 'shows facial gestures or expressions', but was also present for more global target items such as 'ability to maintain student attention'. In other words, Murray & Lawrence (1980) demonstrated that university teaching can be significantly improved through training which focuses on specific classroom behaviours identified in other research as contributing to perceived teaching effectiveness.

Teaching vs. research?

There has been frequent heated conjecture as to whether the relationship between being a good teacher and a good researcher is essentially positive, zero, or negative, and proponents of each view can be heard in faculty club conversations. Several studies, using a variety of criteria, suggest that measures of teaching and research dimensions are essentially orthogonal (Hayes, 1971; Linsky & Strauss, 1975; Rushton *et al.*, 1983).

In the Rushton *et al.* (1983) paper, two separate studies were reported: the first based on an in-depth sample of psychology professors from The University of Western Ontario ($n=52$), the second on a mail survey of psychologists at nine other Canadian universities ($n=69$). In both studies factor analyses were carried out on 29 personality traits, with research and teaching effectiveness composites (which intercorrelated zero) targeted as separate orthogonal factors. Coefficients of congruence calculated between the factor plots obtained in the two studies were found to be high (0·64 for the research factor, and 0·74 for the teaching factor). As can be seen in Fig. 2, even the personality traits that contribute to being a good researcher and to being a good teacher are independent. The only variables loading positively on both dimensions are intelligence and leadership, while meekness suggests being poor on both.

Final comments

In this paper, we have selectively reviewed a sampling of the vast literature on teaching effectiveness in North American universities. More thorough coverage is provided by Marsh (1983), McKeachie (1979) and Murray (1980). However, it should be apparent that a variety of research projects can be undertaken in this important area without inordin-

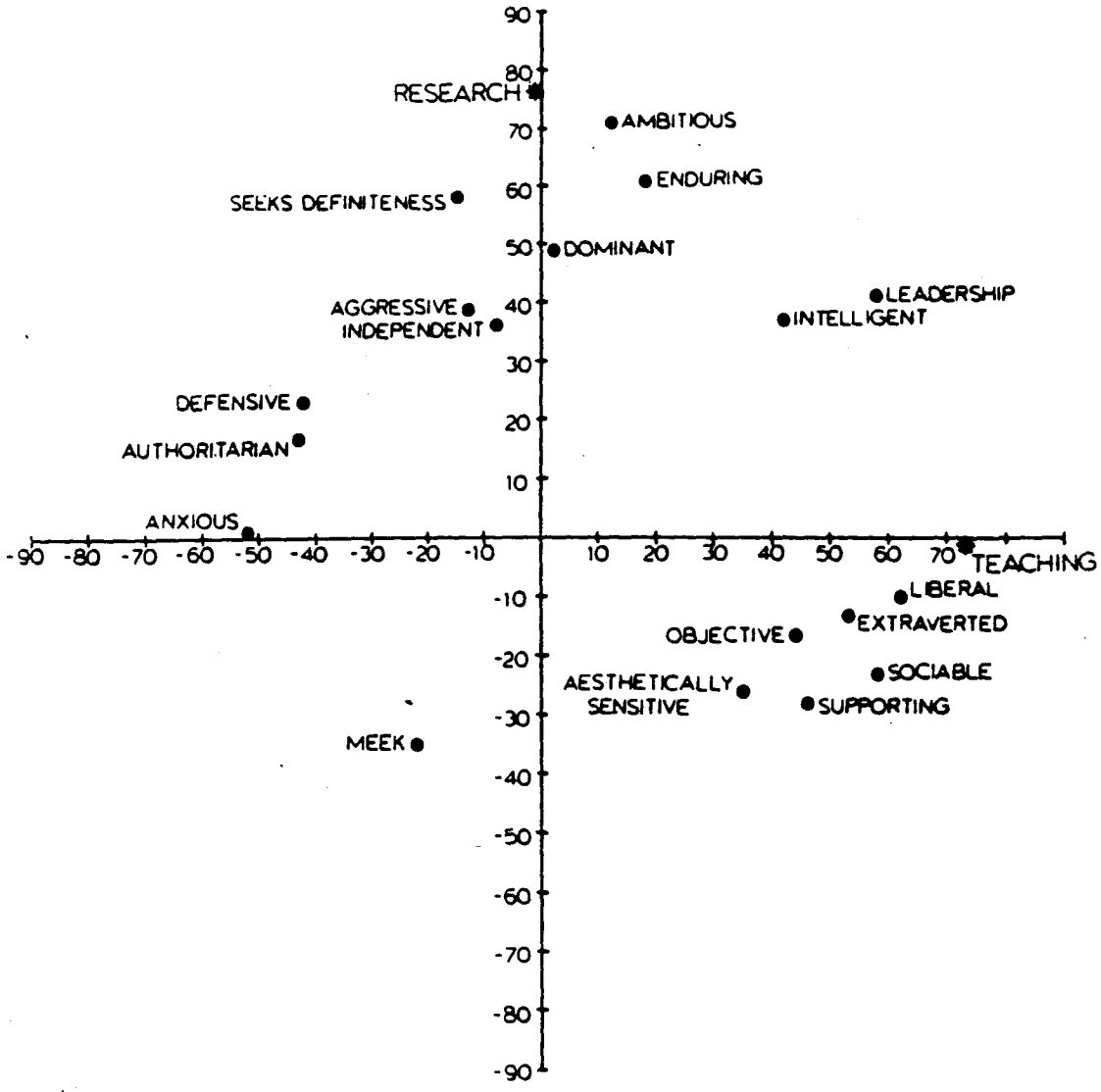


Figure 2. Plot of mean factor pattern coefficients of personality traits on dimensions of research creativity and teaching effectiveness, averaged across two studies. Only those traits with absolute loadings of greater than 0·30 on either factor in both studies are shown. (Source: Rushton *et al.*, 1983).

ate difficulty or effort. Perhaps Thorndike's aphorism pertains here, as it also does when consideration is given to evaluating scientific excellence and research productivity (Jackson & Rushton, in press; Rushton, 1984), namely: 'Everything which exists, exists in some quantity and can, therefore, be measured'. It would be of great interest to see if teaching evaluation results discovered in the North American context are generalizable to their rather different situation in Britain.

References

Carrick-Smith, L. (1985). Assessment of competence in education. *Bulletin of The British Psychological Society*, **38**, 157 (letter).

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, **51**, 281–309.

Erdle, S., Murray, H. G. & Rushton, J. P. (1985). Personality, classroom behavior, and student ratings of college teaching effectiveness: A path analysis. *Journal of Educational Psychology*, **77**, 394–407.

Hayes, J. R. (1971). Research, teaching, and faculty fate. *Science*, **172**, 227–230.

Jackson, D. N. & Rushton, J. P. (eds) (in press). *Scientific Excellence: Origins and Assessment*. Beverly Hills, CA: Sage.

Linsky, A. S. & Strauss, M. A. (1975). Student evaluations, research productivity, and eminence of college faculty. *Journal of Higher Education*, **46**, 89–102.

Marsh, H. W. (1980). The influence of student, course, and instructor characteristics on evaluations of university teaching. *American Educational Research Journal*, **17**, 219–237.

McKeachie, W. J. (1979). Student ratings of faculty: A reprise. *Academe*, **62**, 384–397.

Murray, H. G. (1980). *Evaluating University Teaching: A Review of Research*. Toronto, Ontario: Confederation of University Faculty Associations.

Murray, H. G. (1984). The impact of formative and summative evaluation of teaching in North American universities. *Assessment and Evaluation in Higher Education*, **9**, 117–132.

Murray, H. G. & Lawrence, C. P. (1980). Speech and drama training for lecturers as a means of improving university teaching. *Research in Higher Education*, **13**, 73–90.

Rose, D & Radford, J. (eds) (1984). *Teaching Psychology: Information and Resources*. Leicester: The British Psychological Society.

Rushton, J. P. (1984). Evaluating research eminence in psychology: The construct validity of citation counts. *Bulletin of The British Psychological Society*, **37**, 33–36.

Rushton, J. P., Murray, H. G. & Paunonen, S. V. (1983). Personality, research creativity, and teaching effectiveness in university professors. *Scientometrics*, **5**, 93–116.

Requests for reprints should be addressed to either **J. Philippe Rushton** or **Harry G. Murray**, Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2, Canada.